Psychological Review

EDITED BY
CARROLL C. PRATT, PRINCETON UNIVERSITY

VOLUME 60, 1953

PUBLISHED BIMONTHLY
BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND 1333 SIXTEENTH ST. N. W., WASHINGTON 6, D. C.

Entered as second-class matter July 13, 1897, at the post-office at Lancaster, Pa., under Act of Congress of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized Jan. 8, 1948

LANCASTER PRESS, INC., LANCASTER, PA.

CONTENTS OF VOLUME 60

Vol. 60, No. 1 January, 1953

THE PSYCHOLOGICAL REVIEW

LOGICAL CONSTRUCTS AND PSYCHOLOGICAL THEORY

F. H. GEORGE

Bristol University

This article will attempt to clarify some points which arise in that branch of behavior theory sometimes known as the "theory of learning." Since this paper is mainly concentrated on methodological issues, rather than purely psychological ones, the discussion of the psychological problems will be in general terms, and in particular it will be assumed that the problem of "learning" can be reduced to the problem of "reinforcement." There exist various theories of reinforcement and those of Hull and Tolman will be regarded as representative. It is hoped to show that their differences are largely a result of different methods of theory construction. The contention is that one large part of the problem of "learning" theory can be clarified by a double analysis of existing theories, one from a semantic viewpoint and the other involving a consideration of the place of logical constructs. Also overlapping the semantic question and greatly relevant to the theory builder is the discussion between operationism and pragmatic realism.

In fact, the general problem is that of model making, which in turn leads to a comparison of "molar" and "molecular" analysis. We hope to show, indeed, that the supplementing of "molar" analysis by "molecular" analysis is necessary for the solution of the problems of reinforcement. ("Molecular" will be taken in this article in the approximate sense of Hull to mean neurological, and "molar" to mean "psychological" (12).)

The Problem of Reinforcement

Theories of reinforcement, such as an "expectancy" theory of Tolman or an "effect" theory of Hull, may be investigated both with respect to their differences and their similarities. The fact is apparent that these two theories represent generalizations from different sets of experiments which in all probability exhibit different aspects of behavior. Thus one is tempted to say that the two principles of expectancy and effect are both lacking in generality and refer to "learning" on different levels of abstraction.

There has been a tendency recently for the Hull and Tolman theories to approach an agreement (6, 7, 19, 24) which may be connected with the mixed use of operational and realistic models. There is, however, still present some measure of disagreement which is the direct result of ambiguity involved in the use of logical constructs and in the use of ordinary language.

Meehl and MacCorquodale (15) have suggested that the ultimate nonverbal differences between Hull and Tolman at this stage of development hinge on the centralist-peripheralist problem and may be made to depend on the meaning to be attached to the word "response." These differences can be traced to different historical backgrounds of the two

theories and thus to the acceptance of different philosophical directives for model building.

It is clear that theories may be inadequate for various reasons: (*a*) They may be false as a model of empirical events. (*b*) They may be too specific in view of the dimensions of their assumptions. (*c*) They may be too general and allow insufficient prediction. (*d*) Differences may arise between theories because of their use of apparently incompatible models, and in their confused dimensions and degrees of approximation. (*e*) There may be semantic confusion in the definitions and orders of abstraction, etc. (*f*) The philosophical directives used may be different for different models, although the directives used may not be explicitly stated. Among these reasons, of which one or more may apply to any theories, Hull and Tolman may seem to disagree in (*b*) and (*c*). Hull's degree of definiteness may be at variance with the dimensions of his assumptions and Tolman's relative vagueness fails to give the necessary predictability. But it is in (*d*) and especially (*e*) and (*f*) that we are interested.

Before the discussion of the disagreements, it may be worth while to say a word on the latest Hull revision (7) and its new measure of agreement. It is largely in the clear-cut recognition that performance and "learning" are not identifiable that Hull has moved toward a compromise with Tolman. The price of identification still adhered to by some theorists in essence is to facilitate theory operationally at the expense of arbitrariness. Hull now makes a clear-cut distinction between $_{S}H_{R}$ and $_{S}E_{R}$. This paves the way for the view that reinforcement affects $_{S}E_{R}$ and not $_{S}H_{R}$, and that performance is reinforced independently of "learning." The theory can then be made to allow for sudden changes in performance, which had previously been better accommodated by the Tolman constructs involving "latent learning," etc. This step of agreement is involved because of the implicit acceptance of a mixed operational-realistic model in the place of a purely operational approach. The next step may well be from the other side in an attempt to give the Tolman constructs something approaching operational definitions. The above is intended to suggest that in model building the use of operationism and pragmatic realism are not mutually exclusive, and that it is probable that operationism should be regarded as a borderline or limiting test to an otherwise realist model (1, 2, 4, 5, 8, 20, 21).

Most of the remaining differences between Hull and Tolman seem to fall on definitions and interpretations of those definitions. In particular they appear to surround the ambiguity of words such as "reinforcement," "reinforcer," "response," etc. Also we can, as Seward (19) has done, reduce the differences in one aspect to those of constructs like "habit formation" on the one hand and "mobilization of demand" on the other. The conclusion that the constructs used are ambiguous and may be interpreted in a variety of different ways seems inescapable, and the tracing back of definitions leads to different assumptions dictated by different a priori directives for theory building. These points it is now hoped to clarify in turn.

Semantics, Theory Building and the Use of Logical Constructs

The above brief statement of differences over the theory of reinforcement leads us to reconsider the problem of theory building in psychology, which in essence is the same as that in all other sciences. We have a theoretical model which is of course an abstraction from observations of empirical events and is couched in verbal terms

(sometimes mathematical). The first essential point is the fact that communicable theories in linguistic form involve abstracting. As Korzybski puts it (9), the "word" is not the "object," or perhaps more generally, our theoretical edifice or general model is not the empirical world. We must clearly distinguish between "designata" (nonverbal observable patterns or events), statements about designata, statements about statements, and so on. We cannot assume that the mirroring of designata on our levels of language is unique and unambiguous or that it sets up in the lowest level an isomorphism between language and nonverbal events.

It is of first importance not to identify models and empirical events nor to assume that a model is adequate to all the variables of the empirical events.

The above analysis is also independent of the question whether the lowest level of abstraction in the hierarchy be defined as observations actually carried out (operationism) or as the empirical world which is to be observed (realism). The fact remains that theory building is verbal, and there are many words that are ambiguous if considered out of specific context. These equivocal words may be defined in terms of other words, but even then we are ultimately led back (a difficult process in practice) to a set of undefined or ostensively defined terms which represent the inevitable initial assumptions of our theoretical system—assumptions that are ultimately guided by a priori philosophical directives for theory building.

The above suggests that apart from a haziness involved in the carrying out of operations due to some general principle of indeterminacy, there is also an arbitrariness and/or haziness right at the foundations of our theory. Awareness of the limitations of our model, awareness of the fact that all scientific theory involves approximation is the first step towards offsetting this basic difficulty—"consciousness of abstracting" is Korzybski's term. Korzybski's own model (9, Chaps. 25 and 26), called the structural differential, is worth close attention, but in the form given it seems to represent philosophical realism only. We have followed Korzybski's general rule of using quotation marks about words which appear to be obviously indeterminate in meaning. Thus as Marx (14) has pointed out, there is no precise meaning for words such as "perception," "learning," etc.

We are now led to a consideration of "logical constructs." We accept the differentiation of MacCorquodale and Meehl (13) with respect to intervening variables and logical constructs with the warning that there is of course no clear-cut point of distinction between one and the other. Further we follow Marx in his proposition that constructs act as platforms of a temporary nature which ford gaps in the theory and cloak observable which must ultimately replace these constructs. Now since we do not know precisely what sort of river the logical construct is bridging, there is implicit in its presence a vagueness called "surplus meaning" wherein lie both its strength and its weakness. There must be here an ambiguity which in a general way will vary as to the size of the gap to be filled. A great deal of the vagueness is overcome or minimized when the constructs are used in conjunction with operational definitions. In the sense in which Hull and Tolman use constructs this is not always possible and we shall discuss this point in the last section.

The result of our criticism of logical constructs is to place them, at least when used nonoperationally (i.e., realistically), in a category of "many-meaninged terms" and of ostensively defined

or nondefined basic terms of our verbal system. At every one of these points intervene private and personal usage and looseness of fit engendered by a priori philosophical directives: (*a*) mechanistic, (*b*) teleological, etc.

One further point is the well-known difficulty of mixed models. In physics, for example, the electron may be regarded as particle or wave, but cannot be regarded as completely one or the other. This is quite satisfactory as long as one is "conscious of abstracting" and is using a different model for different dimensions. In psychology, on the other hand, teleological models and mechanistic models, even realist and operational models, tend to be regarded as mutually exclusive. This seems quite incorrect and would appear to be the result of lack of knowledge and understanding of theory building.

A Way Out

The foregoing analysis presents something of an impasse. The claim is that our models are confused in some respects, particularly with respect to their levels of abstraction and dimensions; that there is a degree of ambiguity arising in at least three different levels due to the use of "many-meaninged terms," "logical constructs," and to the inevitable existence of "undefined terms." Much of the vagueness has been removed from such subjects as physics due to the use of the relatively precise language of mathematics and to the consciousness of dimensions. The high degree of vagueness could be diminished in psychology, even in terms of our verbal models, if we were able to state more precisely the existing variables and also increase the number of observables. Thus the continual replacement of logical constructs is necessary; and while many may be transformed into operationally valid intervening variables, the ability to do this, which is involved anyway in the ultimate replacement of constructs, must depend on the increase in the number of observables. Further, since the bulk of constructs in learning theory appear to refer to "organic states," this appears to constitute the clearest evidence for increased concentration on "molecular" levels of analysis.

In the above context, the constructs such as "cognitive maps," $_{S}E_{R}$, $_{S}H_{R}$, are obvious examples. They are necessitated by the distinction between "performance" and "learning." The process of reduction to intervening variables by operational definition or to complete replacement in both cases implies an increase in the number of variables, and thus in the degree of observability.

Apart from this immediate aspect, it is clear that "molar" analysis by itself is quite insufficient for the adequate analysis of the bulk of behavior problems. It is true that it is a vital part, as "molecular" analysis alone may be inadequate. But it seems certain that there must also remain in "molar" analysis alone a degree of approximation, uncertainty, etc., altogether too crude to answer many of the vital questions demanded of behavior theory.

"Molar" analysis on its own seems to contribute to the vagueness inherent in the use of logical constructs. It is too much like trying to use 6-ft. planks to bridge 24-ft. gaps. The cure for this, in part, must, of course, depend on technological advances such as improved instruments of observation, and also the extension of psychological experiments to "molecular" analysis on a larger scale.

The above propositions suggest that the answer to Postman's three questions (16):

(1) What is the agent responsible for reinforcement?
(2) What is it that is reinforced?
(3) What is the basic mechanism of reinforcement?

may begin with question (3) rather than (1) and (2) as is suggested by Seward's recent analysis (19). Thus Wolpe's (27, 28) approach, although perhaps oversimplified, has much to recommend it.

"Consciousness of abstraction," appreciation of dimensions and the nature of approximations and degrees of errors, coupled with a sound knowledge of system building, would seem to constitute the greatest safeguard in the effort to overcome the semantic problems. These will also be simplified, however, with the integration of further levels of abstraction and isolation of variables, which will perhaps allow the use of mathematical language for a great many of the basic problems of behavior theory. The realization that there are points for the inclusion of personal tastes in present theory building may go some way to stop arguments between groups whose views are complementary and capable of synthesis, rather than mutually exclusive.

The next step would seem to involve considerably greater attention to general system-theory; for clearly the theory as a whole, at all levels of science, will affect the parts as well as the whole. The point is illustrated here by the claim of Von Bertalanffy (25, 26) that "teleology" is a formulation derived from a closed thermodynamical system which when used in an open system becomes devoid of meaning. A somewhat similar problem over the word "purpose" arose in the discussion between Taylor (22, 23) and Rosenbleuth and Wiener (18). This last discussion illustrates well the one particular form of semantic confusion that may, if unclarified, vitiate scientific theory. The clarifying points made by Rosenbleuth and Wiener are particularly instructive, especially the fact that degree of causality implies a form of analysis, i.e., statistics and probability theory, and that the use of a term like "purpose" arises as an artifact of abstraction and depends on the characteristics of the model as a whole.

Krech's recent analysis (10, 11) seems particularly fruitful in view of what has been said. The construct "dynamic system" is rooted in neurology and is stated in a form favoring the necessary reducibility—ultimately, of course, to endocrinology, colloidal chemistry, etc.

Two further points may be mentioned. Pratt's analysis of psychological theory (17, 20) points out that all science has the same subject matter. This view is implicit in the foregoing analysis and the truth of Pratt's contention is a starting point for an extension of all scientific theory to the level of general system theory on the one hand, and to the differentiating detail of type of model and dimensions of our particular areas on the other. The second awkward point of testability (3) or verifiability, etc. has been omitted from this note; the belief here is that the awkwardness is largely removed by semantic analysis; otherwise it goes beyond the scope of this paper.

All that has been said in this paper is in essence programmatic. It is hoped to build more positively and specifically upon this analysis in a later paper. It seems to the writer that a program of "integration of science" is more urgent at this time than a greater specificity with respect to particular models.

REFERENCES

1. Bridgman, P. W. *The logic of modern physics.* New York: Macmillan, 1927.
2. Bridgman, P. W. The nature of some of our physical concepts, I, II, and III. *Brit. J. Phil. Sci.*, 1951, 2, 5–7.
3. Carnap, R. Testability and meaning. *Phil. Sci.*, 1936, 3, 419–471; 1937, 4, 1–40.
4. Dingle, H. A theory of measurement. *Brit. J. Phil. Sci.*, 1950, 1, 5–26.
5. Feigl, H. Existential hypotheses. *Phil. Sci.*, 1950, 17, 35–62.

6. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
7. Hull, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.,* 1949, 57, 173–180.
8. Israel, H., & Goldstein, B. Operationism in psychology. *Psychol. Rev.,* 1944, 51, 177–188.
9. Korzybski, A. *Science and sanity.* Edition II. Lancaster, Penna.: Science Press, 1941.
10. Krech, D. Dynamic systems, psychological fields, and hypothetical constructs. *Psychol. Rev.,* 1950, 57, 283–290.
11. Krech, D. Dynamic systems as open neurological systems. *Psychol. Rev.,* 1950, 57, 345–361.
12. Littman, R. A., & Rosen, E. Molar and molecular. *Psychol. Rev.,* 1950, 57, 58–65.
13. MacCorquodale, K., & Meehl, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.,* 1948, 55, 95–107.
14. Marx, M. H. Intervening variable or hypothetical construct? *Psychol. Rev.,* 1951, 58, 235–247.
15. Meehl, P. E., & MacCorquodale, K. Some methodological comments concerning expectancy theory. *Psychol. Rev.,* 1951, 58, 230–233.
16. Postman, L. The history and present status of the law of effect. *Psychol. Bull.,* 1947, 44, 489–563.
17. Pratt, C. C. *The logic of modern psychology.* New York: Macmillan, 1939.
18. Rosenbleuth, A., & Wiener, N. Purposeful and non-purposeful behavior. *Phil. Sci.,* 1950, 17, 318–326.
19. Seward, J. P. Secondary reinforcement as tertiary motivation—a revision of Hull's revision. *Psychol. Rev.,* 1947, 54, 1–8.
20. Symposium on Operationism. *Psychol. Rev.,* 1945, 52, 241–294.
21. Symposium on Existential Hypotheses. *Phil. Sci.,* 1950, 17, 2.
22. Taylor, R. Comments on a mechanistic interpretation of purposefulness. *Phil. Sci.,* 1950, 17, 310–317.
23. Taylor, R. Purposeful and non-purposeful behaviour: a rejoinder. *Phil. Sci.,* 1950, 17, 327–332.
24. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.,* 1948, 55, 189–208.
25. von Bertalanffy, L. An outline of general system theory. *B. J. Phil. Sci.,* 1950, 1, 134–165.
26. von Bertalanffy, L. The theory of open systems in physics and biology. *Science,* 1950, III, 23–29.
27. Wolpe, J. An interpretation of the effects of stimuli (patterns) based on current neurophysiology. *Psychol. Rev.,* 1949, 56, 284–291.
28. Wolpe, J. Need-reduction, drive-reduction and reinforcement—a neurophysiological view. *Psychol. Rev.,* 1950, 57, 19–26.

[MS. received October 15, 1951]

Psychological Review
Vol. 60, No. 1, 1953

PHYSICAL PSYCHOLOGY

R. C. DAVIS

Indiana University

Predominant Traits of Systematic Psychology

The dominant feature in psychological theories during the last two decades seems to have been a concern for logical precision, with the object of fashioning a structure like physics. As physics has its logical constructs, so psychology has developed a great variety of hypothetical concepts and intervening variables, generated in this case by the psychologists' efforts to organize behavior or "experiential" data, and having no other necessary reference. In some cases even the language describing the data is of a sort peculiar to psychology. Much research has been guided by this kind of thinking, though not all of it. Some deals only with physically defined variables. It is the purpose of this paper to explore the rationale for this latter kind of work and see whether it might be generalized. To designate this sort of thinking conveniently the term "physical psychology" is suggested.[1]

This situation in psychological theory seems to have come about because of the severe criticism during the 1920's of the uninhibited "neurologizing" which was then current. In an excess of caution, many writers undertook to construct a pure science of behavior as though no other knowledge existed. So we have terms like drive (Hull, 5), psychic force (Lewin, 8), cognitive maps (Tolman, 16), conflict (Miller, 10), etc., terms which refer only to the class of data they were constructed to explain. Somewhere in the domain of science, no doubt, there must be a choice of nonreducible terms, but it is not necessary that each particular science should make a new choice. The semantic clarity and sophistication of the enterprise are, of course, a great merit of the recent psychological writing. But the application of scientific methodology to the subject *de novo*, as it were, is a procedure which has perhaps already led us as far as it can. Instead of seeking the ideal of a science *like* physics, we might do well to try making psychology a *part of* physics, or more precisely, a part of the body of general science which regards reducibility as an obvious requirement of a good theory.

With as much elbow room as nonreducibility allows it is not surprising that theorists can execute a wide variety of maneuvers. But there arises a suspicion, at least, that there is no unique solution when the theorist is allowed such an open field. Paradoxically theories are too easy to make. If a theorist, having taken a few peeks at the data, selects a set of hypothetical constructs and uses them in his equation (mathematical or verbal) which predicts the data, he is restricting himself somewhat, but probably not enough to exclude alternatives. Another set of equations with different terms might accord with the same facts, or facts that are so nearly the same as to be nondiscriminable. If this suspicion is correct, there will be no "crucial" experiments within the sphere of behavior to decide between a good number of theories.[2] Each

[1] Logical empiricists have, of course, recommended "physicalism" for some time, but they have not gone so far as the position here presented. Probably they would not insist upon the kind of reduction here advocated. Therefore, it seems advisable to distinguish between a physical and a physicalist psychology.

[2] Meehl and MacCorquodale (9) find this is true of the Hull and Tolman theories.

one will be able to assimilate any possible facts.

For one who is committed to an independent behavioral science there remains only an appeal to parsimony or elegance. Although theories are usually believed by their advocates to have these qualities, an explicit demonstration of them is uncommon. The principle of simplicity is in fact not very simple to apply. It is not very clear whether one should try to minimize constructs, assumptions, propositions, or all of these. In any case, to apply the principle of parsimony, a definite field of observations needs to be marked out; for if one is to compare two explanations, it must be certain that they cover the same territory. In general, one theory may be expected to operate more conveniently in one part of a field, another in a different part. To evaluate them one needs to decide which part is more important and how much awkwardness here is equal to how much awkwardness there. It is no wonder that psychologists shrink from such a comparison. Parsimony is doubtless needed as a last resort in the philosophy of science, but it does not offer an attractive short cut.

The Possibility of Physically Restricted Theory

Conceivably, a unification of theory could be achieved by fiat if writers would agree to one particular set of logical fictions. At periods in the history of psychology, something of the sort seems to have happened; periods which have been the dullest and least profitable of all. For scientific progress, it seems one must be able to consider alternatives and have some criterion for deciding between them.

The criterion here proposed is the observability of the thing hypothesized instead of merely the observability of the behavioral consequences. Only those constructs would be tolerated which could be reduced to those of physical science. To illustrate: "Hunger drive" defined as an influence on behavior generated by lack of food is a construct which is not restricted by such a rule (Spence, 14). On the other hand, a blood sugar deficiency as a possible connecting link between food deprivation and certain behavior is a construct of the second kind. (Although constructs would be reducible to the ultimate ones of physics, they need not, of course, be so reduced on all occasions.)

Such a rule narrows the area in which a theorist may grope, and in so doing, confers a number of benefits. With it psychology can form junctions with the underlying sciences. In modern times, chemistry, geology, biology, and physiology, for example, have for the most part restricted themselves to explanatory constructs of the "reducible" sort. They have, as it were, extended the domain of physics into their fields; or to put it better, they have had a good deal of success constructing one general science with the same language throughout. In an earlier time they did not always proceed so, but now constructs like entelechy, valence, and central excitatory state have vanished or acquired a "physical" meaning. On the other hand, in the social sciences, the use of "free" constructs seems to be rather prevalent. In these, as one would expect, there is not much connection between one science and another, nor even sometimes between one topic and another in the same science.[8] So, if psy-

[8] To be sure, there have been complaints that the biological sciences, for instance, are wanting in general unifying principle, and, therefore, fail to be truly scientific. A "system of physiology" would, indeed, now fall upon the ear like an echo from antiquity. Yet blundering about as they may be in comparison to mathematics or physics, these empirical sciences nevertheless have their triumphs and are hardly to be despised as models of con-

chological theorists can work under the restraining rule, there is a chance that the field of general science will spread into the realm of psychology, while lines of communication with the older provinces are kept open.

One way to reach this Utopia of unified science would be to start with a likely set of "behavioral" constructs, by hypothesis give them an independent meaning, and proceed to experimental test. A behavioral theorist would postulate a variable which has such and such a relation to behavior, to the stimulus, or to both, and thereby set the problem for the physiological experimeter of discovering, somewhere in the organism presumably, an actual condition or event which has the stated relations. An enterprise of that sort has often been favored; it seems to be the hope of some "behavioral" theorists (Hull), and even physiological psychologists (11), (3) appear content to take their framework from behavioral theory and try to fill it with physiological referents. Krech (7) complains that the theorist in such a proceeding gives the experimenter no clue about where to look. He may also send the experimenter on a fool's errand. For the postulated construct may belong to a set which is only one of a great number of alternatives, only one of which could have a physiological meaning. By good fortune the experimenter might hit upon the right formulation. But to proceed so is to give up the guidance extra-behavioral knowledge may offer in the original choice of concepts: a guidance which we badly need if the preceding remarks are correct. To put that knowledge out of mind in psychological theorizing while hoping for later independent verification seems rather like a game of darts played blindfold. Later inspection may show the aim was good, but why play that way anyhow?

Instead, the originator of explanatory concepts can be guided by the rule that they should have an operational meaning outside the realm of behavior. The theorist, in other words, would construct a hypothesis with a much broader set of facts in mind; indeed all the facts that are represented by the "physical" constructs he would use as the behavioral events he is theorizing about. Of necessity his freedom is reduced, as the foregoing argument would have it. Another of the effects would be to place physiological and behavioral psychology on the same footing: physiological psychologists would no longer strain their wits to find a physical lineage for constructs of pure Olympian birth. Perception, emotion, motivation, habit strength, expectancy, and other terms which offer names for ignorance (or maybe a trifle more) would probably disappear.

Other criticisms of these concepts could be made: It may be difficult to make such terms very definite by any means at all. But, since they are not meaningful in a physical sense, they would fail to meet this canon for explanatory concepts, however well specified in some other way.

Terms which would fit the proposed restriction would surely be physiological in good part, but not altogether so. Physiological terms would be needed in place of properties of the individual (permanent and temporary) that are mentioned or implied in so many theories, and in place of hypothetical processes that take place between stimulation and response. Theories might refer, for example, to oxygen level or hormone content of the blood, neural conduction characteristics, or state of contraction of such and such muscles, as intervening between a certain stimu-

duct. If they have a unifying principle, it is the self-denying principle of using only general scientific concepts.

lus and its response, or determining one or another kind of response to a stimulus. (There is no reason, of course, why theories should be restricted to the nervous system.)

But the psychologist also needs to describe external factors that influence behavior. His task as a whole would be simply to uncover the chain of events which leads to a given action. The immediate precursors will be within the organism, of course, but the chain will certainly extend to events in the world outside. It is likely, further, that the chain will be found to cross the border of the skin repeatedly if it is traced far enough. These external events may be called stimuli for convenience (though the name seems rather misleading), but in any case there would be no reason for demarking them very sharply from the internal events. Both would be treated the same way. The physical psychologist would need to make hypotheses about these external events as well as about those inside, and here the appropriate concepts would be either immediately physical or reducible to physical through the inorganic sciences.

The outside world seems such a solid thing that some theorists feel that they can lay all their burdens on it. But when overlaid with psychological portmanteaus like valences, psychic distances, stimulus functions, and sign-gestalts, the solid terrain seems to vanish beneath the waves and leave the baggage floating. These abstractions are derived, evidently, from observations on behavior in the same manner as those which seem to refer to the organism, and they suffer from the same indeterminacy. For a "stimulus" (external event) to qualify under the proposed canon, it would have to be something which an experimenter could ascertain without there being any organism for it to work on.

Behavior Description under a Reductive Rule

A physical description of the *data* that are the starting point of psychological theory would evidently lay the groundwork better for a physically bound *theory*. If the latter is a proper goal, as has been argued, the first step should be in that direction too. But unfortunately a great deal of psychological data is not given in physical terms, even when it might be; and some theorists expressly defend the use of nonrestricted terms. Cantril (2) proposes that any sort of descriptive terminology the psychologist can think of will be a good basis for beginning an investigation. If nonphysical terms are to be excluded, two questions need answering: (*a*) what sort of description would result from the application of the physical criterion? and (*b*) what, if anything, would be lost by the abandonment of the nonphysical descriptive terms?

In a sense the physical specification of data is unavoidable. The events that constitute the subject matter of the science, it is assumed, are the actions of individuals; that is, their movements and their utterances, and each single event has its physical characteristics whether noted or not. Even though a psychologist should say he is describing subjective experiences, he is using the utterances and sometimes the actions of his observers as indicators, and these have their physical characteristics. Such characteristics are available then in all data, but they are used in different ways in the next stage of the scientific process.

Because no events apparently are the exact replicas of others, the investigator must form them into groups on the basis of certain resemblances before he can continue. It is in this classification that psychologists differ in their usage and there seem to be three principal

bases on which a classification is carried out.

1. The classification may be explicitly based on some physical characteristic of the activity which is the response event. The plan would be, then, to study the class of events marked by a forearm flexion, for example, or the exertion of a given force, or the occurrence of a certain pattern of muscular or neural excitation. The consequences of the events would be disregarded in the course of inquiry into the origins. The physical nature of the definition is obvious in this case. It seems to be the kind of procedure that Guthrie (4) has recommended.

2. No less dependent on physical specification is a plan of classification which would group together those events which have the same effect on objects around the individual or on the object-individual relationship. On this basis all actions resulting in, say, bar pressing, or the transportation of goods to a given point, or the approximation of an individual and an object would be treated as unit classes for the purpose of investigating their determiners. This kind of classification has been favored by Skinner (13) and Nissen (12). As they point out, the specifications here are physical also, although they refer to effects outside the organism.

3. A third manner of classification depends on the experimenter's response to the subject's behavior, a response made without explicit rules. The subject makes a response and the experimenter (or another observer) judges that the response is, for example, aggressive, or submissive, or insightful, or possesses any other adverbial quality that happens to be of interest to the experimenter. The experimenter, that is to say, makes a verbal response to the subject's behavior and then considers all behavior which provokes the same verbal response from the experimenter as belonging to the same class. It is often thought that in doing so, he is defining classes which have no common physical characteristics. But the common physical event is present nevertheless; it is the response of the observer, and behavioral events which have similar consequences in the observer's behavior are assimilated to the same class.

Though all three methods are dependent upon a physical characteristic, there is a question whether they all depend upon a physical characteristic of the *subject's response*. About the first there is no question since it is expressly based upon such characteristics. The second is a little further removed. Whether it would be a strategic beginning very likely depends on whether the state of external affairs has a feedback effect so that it continues to act as a stimulus until a particular end result is reached. To what extent it does can be found out, of course, only by experimentation, for there seem to be examples of responses so controlled and not so controlled, and it would be very hazardous to take for granted the universality of either kind of response mechanism. It would be a handicap, furthermore, to have the response specified in terms of external results when they play no role in the regulation of the response.

The third type of response specification starts a long way from the response event. A good many factors other than the subject's response are likely to condition the observer's response. This is partly determined by what he is observing, but also no doubt a good deal by his vocabulary, his predilections, his standards, etc. All such things extraneous to subject's behavior would act as "noise" to corrupt the information which is desired. Possibly there are instances, nevertheless, where such reports of observers need to be used as a starting point. They may be justified on the same ground as the use

of the external consequence according to the second type of approach. That is to say, if the subject persists in a line of action until a certain response is produced in the observer, it would be appropriate to class together subject's responses which lead to the same result in the observer's response. A physical account would begin with the observer's final action and describe in physical terms what can be seen of the actions of both subject and observer leading up to it. In sum, there will be a system with loose couplings in both directions and extra systemic influences entering at both points. Experimental or other analytic procedure would be necessary to uncover what factors in the behavior of the two and what outside factors play the role of stimuli for the two.

No barrier therefore stands in the way of a physical description of all sorts of behavior. The physical aspects are there if we but note them. Nor does it seem that anything would be lost by such tactics except some unprofitable entities named by the many adverbs in the common language.

In the course of statistical treatments of data which are physically defined in the beginning, psychologists may be inclined to forget this humble origin and imagine that some statistical abstractions represent causal factors underlying the observations. "Measured" traits, whether simply test scores or refined to Thurstone factors, sometimes seem to have this specious reality, and to present the theorist with the illusory problem of how to connect such nonphysically defined entities with the realm of general science. The original data in these cases are, of course, the responses people make in taking the tests—responses of putting pencil marks in certain blanks rather than in certain others. These are physically defined data of the second class, which may perhaps be well summarized by the statistical abstractions, but hardly originated or explained by them.

In very much the same case are the dimensional variables that come out of psychophysical research. The sensation scales of various kinds, loudness, brilliance, volume, for example, are based on concepts which are not a part of the common scientific language. It is not possible to reduce any of them to physical units except by means of behavioral data. As Bergmann and Spence (1) observe, they are logical abstractions from behavior data, at bottom just summary descriptions of certain responses. However useful these may be for some purposes, it is still the responses which need to be accounted for, not the "as-if" world which would account for the responses if it had a real existence. If one succumbs to the temptation of reifying terms like threshold, constant error, sensation dimension, he is attributing the physical responses to nonphysical variables.

In hoping for more valid measures of such things, Stevens (15), for example, implies that they exist independent of the responses (yes-no utterances in a discrimination experiment). This is a proposition which surely ought not to be taken for granted. A hunt for the physical parallels of these entities (the "physical dimensions of consciousness") would very likely have the same outcome as a polar expedition to discover St. Nicholas.

The General Effects of a Physical Program

A psychology written entirely in the common scientific language would differ a good deal from general psychology as it is now presented and eliminate much that is now discussed as psychological theory. At the outset the burden of distinguishing psychology from physiology on the basis of source or kind of

data, scale of data, a method or purpose of study would be lifted.

Another burden which psychologists would no longer be obliged to sustain would be the definition of the present main categories of psychology. These chapter headings, such as perception, motivation, emotion, personality, etc., seem troublesome to most writers already, and here and there some of them have been discarded, e.g., by Keller and Schoenfeld (6). No doubt these words were originally intended as names for mental entities, but writers at present would try to avoid such a naive reference. Consequently, at the beginning of each topic a writer or lecturer will say, in effect, "Now I shall speak about ———. But, of course, you are to understand that there really is no such thing as ———." The terms, of course, are not reducible to physical components and have to be treated as logical constructs or self-defining descriptions.

What categories would replace these ancient ones as a basis for organization, it would require a prophet to say. But, in any case the content would be statements about the connections between (*a*) physical conditions in the environment, (*b*) the physiological, physical, and chemical conditions and events in the organism, and (*c*) the physical consequences. It is impossible to answer the question "What category and concept would replace such and such a one now in use?" It is not likely that any new concept would exactly replace an old one; indeed, in such a case, there would be little gain. One may hazard a guess in a particular instance to serve as an illustration. With the abandonment of "motivation" as a category, the construct of "motive" or "drive" defined in the current roundabout fashion would also disappear. Some of the experimental results now handled with these terms could possibly be assembled into a topic "persistent systemic physical and chemical states affecting responses." This would include an account of the effects of nutritional states, hormonal conditions, temperature states, etc., presenting material now included in the discussion of motivation. It would also contain material on the effects of oxygen deficiency, fatigue products, bacterial toxins, etc., which are not generally put into chapters on motivation (or in any other chapters for that matter).

In this exposition, certain material now discussed under the topic of motivation would be out of place. It would hardly include experiments which are concerned with the effect of electric shock on learning, for example. Nor would there be any reason to speak of "conflict" since physical and chemical states do not war with one another. "Ego maintenance" is obviously not in the physical universe of discourse. The facts now assembled under these subheads would be put under some other general topics.

Thus, it will be seen, the insistence upon physical science would probably result in a different organization of present material. If the argument here offered is correct, it would, when facts are available, come nearer to producing a definite psychology. Of course, the result for a long time to come would be much less tidy than the systems now in vogue. For with restricted constructions no single insight, however brilliant, is likely to reveal the theory of everything, nor a few simple postulates to imply all the complexities of life. Although there is nothing to prevent speculation from outrunning facts by a great distance, performances of this sort could not but seem rather eccentric. The sprint might too easily be in the wrong direction.

The use of physical concepts in psychology brings to mind earlier proposals such as those offered by the behaviorists

several decades ago. But their reasoning seems to have been different, their position apt to be sustained by little more than an appeal to tough-mindedness. In practice there was too frequently a simple "physicalizing" of old concepts. The terms "instinct" and "emotion," for example, were merely asserted to refer to bodily states. The traditional, vaguely defined categories and processes were taken as the starting point and a likely-seeming physiological process was proposed as their substance. This physiologizing of old psychology is, of course, a different undertaking from the construction of a physically reducible psychology from the beginning. In other words, early behaviorists were not sufficiently radical.

General psychology written according to the physical prescription is something yet to appear. Experiments in these terms are nevertheless not wanting. Many experimental reports seem to have little connection with any current psychological structure, but their implicit object seems to be the kind of behavior analysis here projected. Further, physical information can often be extracted from its bed in alien theory. Indeed, if the foregoing analysis is correct, the experimenter will actually have identified the data by physical specifiations and he will often state them even when he considers them unimportant. The formulation of physical theory for the fields of psychology thus seems possible. Naturally the most persuasive evidence would be the systematic application of it even to a small area where it had not been used before. It seems as though it would work. The argument for the path offered here is basically one of expediency; it is a route which promises more in the long run. And expediency, of course, is relative. Where there is yet no map referring to the Greenwich meridian (or even to any coordinates), stories of the region may not be quite useless—so long as they do not merely instruct the traveler to follow his nose. But the explorer will want to replace them as soon as possible with readings of compass and transit.

REFERENCES

1. Bergmann, G., & Spence, K. W. The logic of psychophysical measurement. *Psychol. Rev.*, 1944, 51, 1–24.
2. Cantril, H., *et al.* Psychology and scientific research: II. Scientific inquiry and scientific method. *Science*, 1949, **110**, 491–497.
3. Freeman, G. L. *Physiological psychology.* New York: Van Nostrand, 1948.
4. Guthrie, E. R. Psychological facts and psychological theory. *Psychol. Bull.*, 1946, 43, 1–20.
5. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
6. Keller, F., & Schoenfeld, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950.
7. Krech, D. Dynamic systems, psychological fields and hypothetical constructs. *Psychol. Rev.*, 1950, 57, 283–290.
8. Lewin, K. The conceptual representation and measurement of psychological forces. *Contr. psychol. Theory*, 1936, 1, No. 4.
9. Meehl, P., & MacCorquodale, K. Some methodological comments concerning expectancy theory. *Psychol. Rev.*, 1951, 58, 230–233.
10. Miller, N. E. Experimental studies in conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944, 431–465.
11. Morgan, C. L., & Stellar, E. *Physiological psychology.* New York: McGraw-Hill, 1950.
12. Nissen, H. W. Description of the learned response in discrimination behavior. *Psychol. Rev.*, 1950, 57, 121–131.
13. Skinner, B. F. The generic nature of the concepts of stimulus and response. *J. gen. Psychol.*, 1931, 12, 40–65.
14. Spence, K. W. The postulates and methods of 'behaviorism.' *Psychol. Rev.*, 1948, 55, 67–78.
15. Stevens, S. S. Mathematics, measurement and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.
16. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, 55, 189–208.

[MS. received October 24, 1951]

Psychological Review
Vol. 60, No. 1, 1953

A GRADIENT THEORY OF MULTIPLE-CHOICE LEARNING

JOHN OLIVER COOK
New York University

Ever since Yerkes (14) modified Hamilton's (2) apparatus for studying learning behavior, the multiple-choice method has provided psychologists with a rich profusion of facts and a theoretical problem. It would be hard to point to an area outside the field of sensory psychology that matches this one for plethora of data and paucity of theory able to account for any considerable share of the data. Yerkes' original view was that multiple-choice learning requires ideation. Hunter (4) dissents from this view. Washburn (12) holds that only the middle-door problem, in which the middle one of, say, three doors is the correct choice, requires ideation, and that the end-door problem involves only a simple position habit. Burtt (1) also takes this stand with regard to the end-door problem, and Spence (9) believes that "strong support may be found for the view."

What is certainly the case is that the middle-door problem is harder than the end-door problem. It is, in fact, harder than the alternate end-door problem and the second-door-from-end problem. To quote Spence's review (9) of experimental results:

> Only one subject, a European siskin[1] (7), has been reported able to solve the middle-door problem. Four chimpanzees (15) failed in this problem, as did one monkey (13) and two pigs (16). The problem next in difficulty, the alternate-end-door, was successfully solved by two pigs (16), one of two monkeys (13), and only one of four chimpanzees. Likewise, only one chimpanzee of four was able to succeed with the second-door-from-end problem (13). This latter was also solved by three monkeys (13 and 10), two pigs, a skunk (5) and a marten (5). It proved too difficult for the orangutan (13) and two white rats (1). All subjects tested, chimpanzee, orang, monkey, pig, cat, white rat, and crow, have been reported as succeeding with the simple end-box problem.

The question remains: Why is the middle-door problem more difficult than the end-door problem? It may be subjectively satisfying to laymen and psychologists alike to say that in the middle-door problem the animal is confused by the doors on either side in a way that it is not confused in the end-door problem, but this skirts the issue. Why is it more confused? It seems similarly inadequate to say that the middle door is "imbedded" to a greater extent than the end door. Indeed, such hypotheses as these are hardly more than photographs of the problem. Underwood (11), though he has no well-developed theory, regards multiple-choice learning as a species of concept formation, but this, of course, does not explain why the middle-door concept is harder to form than the end-door concept. It may be of some help to say, with Washburn and Burtt, that the one involves ideation and the other requires only a simple position habit, but whatever virtues such a dichotomous theory may have, it is certainly less parsimonious than a theory that explains the results on a single basis. Moreover, the theory of Washburn and Burtt runs counter to the widespread suspicion among psychologists that chimpanzees are brighter than siskins.

Another fact about multiple-choice learning is one mentioned by Under-

[1] Defined by *Webster's Unabridged* as: "a small, sharp-billed, chiefly greenish and yellowish finch (*Spinus spinus*) of temperate Europe and Asia, allied to the goldfinch."

wood (11) and also by Munn (6), which is that human subjects find it very difficult to solve a multiple-choice problem when the correct choice alternates about the middle; that is, for example, when the correct choice in a five-choice apparatus is No. 2 on the first trial, No. 4 on the second trial, No. 2 again on the third trial, etc. As far as the present writer knows, no theory has yet been advanced to account for this.

A third phenomenon that needs explaining is the way in which chimpanzees shift from the correct response on one problem to the correct response on a following problem. Spence (9) noted that the animals made the change gradually. If the former correct response had been to the second door from the left end of a six-door setting, and the correct door in the new problem is the second door from the right-end door of seven doors, the chimpanzees, after failing to get reinforcement at the formerly correct door, shift from left to right, trying each door in succession.

A fourth phenomenon of multiple-choice learning, and one also remarked on by Spence (9), is the rather curious fact that in shifting from one correct response to another, all the chimpanzees, without exception, exhibited a tendency to overshoot the mark. For example, in the hypothetical case just mentioned, instead of stopping at door No. 6, the correct door, the animal, after being rewarded at that door, proceeds to door No. 7.

The hypothesis advanced here attempts to account for all four of these phenomena: (*a*) the fact that animals find the middle-door problem harder than the end-door problem, (*b*) the fact that human subjects find it very difficult when the correct choice alternates about the middle, (*c*) the fact that in chimpanzees the shift from a formerly correct response to a new correct response is a gradual approach process, and (*d*) the fact that in so shifting, chimpanzees have a tendency to overshoot the mark and respond to the next door just after being rewarded at the correct door.

Explanations of other phenomena of multiple-choice learning can probably also be deduced from the theory, and one or two of these will be suggested. But the first task is to expound the theory and show that the four previously mentioned phenomena are predictable on the basis of the theory.

The present hypothesis owes an enormous debt on the theoretical side to Spence's account (8) of the transposition experiment and a scarcely smaller debt on the experimental side to Spence's monograph (9) on multiple-choice learning in chimpanzees. Like its parent theory, this one is based on the notion of stimulus generalization, and the two theories have four assumptions in common. They assume that there is an excitatory gradient about a stimulus to which a reinforced response has been made. The second assumption is that an inhibitory gradient is set up about a stimulus when responses to that stimulus are *not* reinforced. The third assumption is that the excitatory gradient has a higher maximum than the inhibitory gradient and also has a wider extent. The fourth assumption is that a response tendency is the algebraic sum of the excitatory and inhibitory potentials.

Given the basic assumptions, let us now hypothesize some numerical values for the gradients. These numerical values may be incorrect, but that does not invalidate the theory. In order to do that they would have to be so grossly wrong as not to constitute gradients at all. Several different sets of numbers no doubt will yield adequate predictions. The only virtue of these particular sets is that they, too, will explain the phenomena. Let us assume that

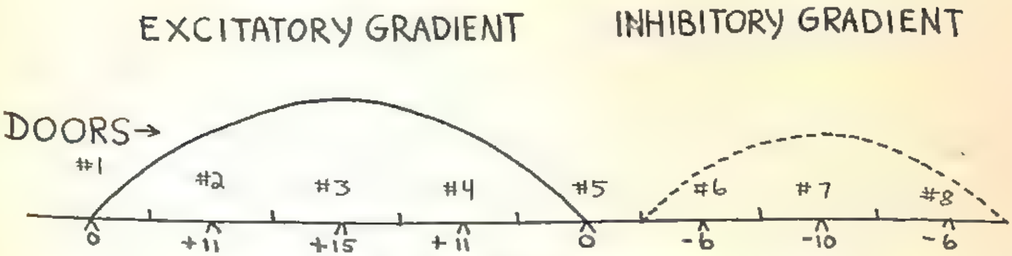


Fig. 1. Numerical values for excitatory and inhibitory gradients

the excitatory gradient has a maximum of 15 over the correct door; at one remove it has a value of 11; and it falls to zero by the time it reaches the door at two removes from the correct choice. The inhibitory gradient is assumed to have a maximum of 10, to fall to 6 at one remove and to zero at two removes. Thus the two gradients would be as pictured in Fig. 1. In computing response tendencies the excitatory gradient is given positive values and the inhibitory gradient negative ones. These values are indicated below the numbered doors. It may be noted that though the excitatory gradient is wider than the inhibitory one, it is not wide enough to have a numerical value at two removes greater than zero.

On the basis of these hypothesized values, the resultant response tendencies for the end-door problem in which door No. 3 is the correct choice would be as depicted in Fig. 2. Thus the correct door enjoys an absolute advantage over its closest competitor (door No. 2) of the difference between − 5 and + 9, or 14 points. The greater the difference between competing response tendencies, of course, the easier the learning.

This difference contrasts sharply with the situation that obtains in the much more difficult middle-door problem as diagrammed in Fig. 3. Here, owing to the overlapping of inhibitory gradients over the correct (middle) choice, the middle door has only the very slight advantage of two points over its competitors.[2] Thus the first phenomenon is accounted for by the theory.

[2] If one is interested in the question of phylogenetic differences in multiple-choice learning, it is just possible that one of the answers may lie in this: that higher organisms are capable of responding to smaller dif-

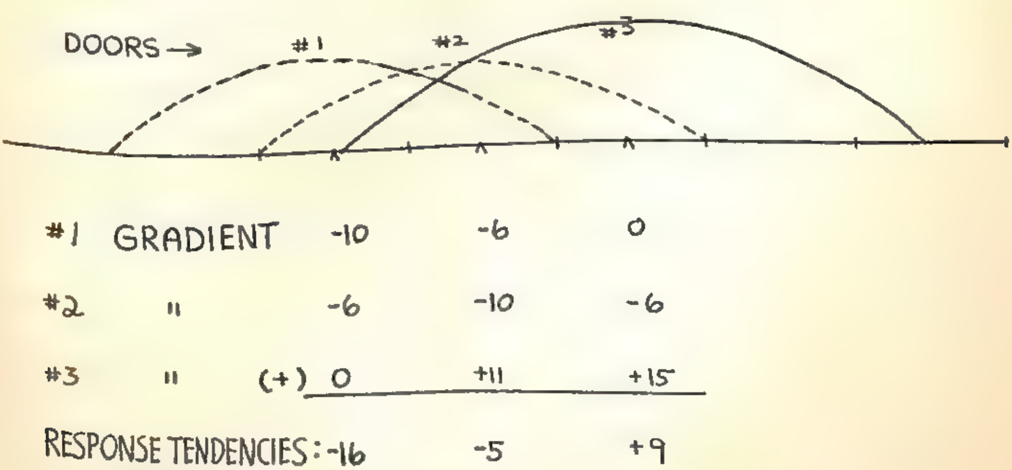
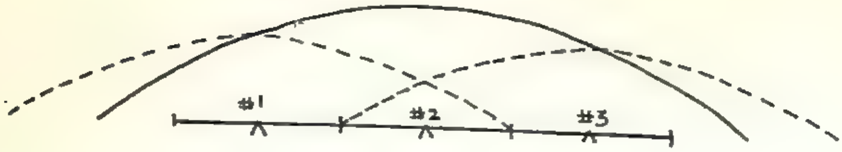


Fig. 2. Response tendencies for end-door problem

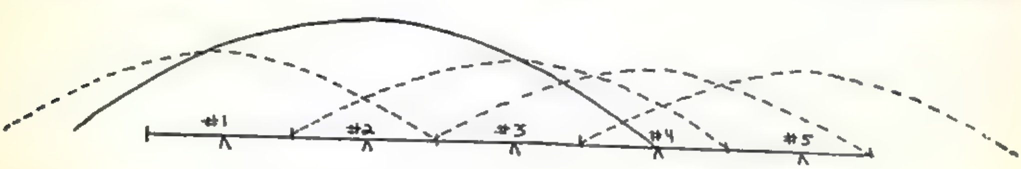


#1 GRADIENT		-10	-6	0
#2	"	(+) +11	+15	+11
#3	"	0	-6	-10
RESPONSE TENDENCIES:		+1	+3	+1

Fig. 3. Response tendencies for middle-door problem

To move on to the second phenomenon to be explained—namely, why human subjects find it so difficult to solve a multiple-choice problem when the correct response alternates from side to side about the middle on successive trials—let us suppose that we have a five-choice apparatus in which the correct choice is No. 2 on the first trial,

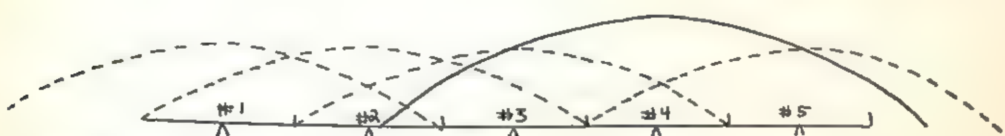
ferentials than are lower forms. The rather puzzling case of the single European siskin will be discussed later.



#1 GRADIENT			-10	-6	0	0	0
#2	"	(+)	+11	+15	+11	0	0
#3	"		0	-6	-10	-6	0
#4	"		0	0	-6	-10	-6
#5	"		0	0	0	-6	-10
RESPONSE TENDENCIES			+1	+3	-5	-22	-16

Fig. 4. Response tendencies after first set of trials

No 4 on the second, No. 2 again on the third, etc. Let us also assume that the subjects try all the choices an equal number of times. We shall thus call the first trial "a set of trials" and compute the response tendencies after the first set of trials as follows (Fig. 4). It is apparent from the response tendencies that the least likely response on the second set of trials will be door No. 4, which is precisely the correct choice. Certain response tendencies have accumulated from the first set of trials. If we compute the response tendencies that accumulate during the second set of trials and add to them the response tendencies that accumulated during the first set, we can then make a prediction about the likelihood of the correct response (to door No. 2 again) on the third set of trials (Fig. 5). Just as the inhibitory potential piled up on door No. 4 during the first set of trials, so it piles up on door No. 2 during the second set with the result that the total response tendencies make a response to door No. 2 (the correct choice) on the third set of trials as unlikely as a response to door No. 4 and more unlikely than a response to any other door. After the third set of trials door No. 4 again becomes the least likely; after the fourth set door No. 2 again becomes equally unlikely with No. 4; and so on ad infinitum. In fact, if it were not for the presence of other factors in the situation, such as the fading of the gradients through forgetting between trials, etc., the only thing that would make this multiple-choice problem soluble at all is the circumstance that the experimental subjects do *not* make each choice equally often. On the first set of trials a subject may make the correct response before he makes all the incorrect responses, and this will eliminate for this set of trials one or more of the inhibitory gradients. On the second set all the different incorrect responses may not occur, thus further reducing



	#1	#2	#3	#4	#5
#1 GRADIENT	-10	-6	0	0	0
#2 "	-6	-10	-6	0	0
#3 "	0	-6	-10	-6	0
#4 " (+)	0	0	+11	+15	+11
#5 "	0	0	0	-6	-10
RESPONSE TENDENCIES:	-16	-22	-5	+3	+1
ACCUMULATED FROM FIRST SET OF TRIALS	+1	+3	-5	-22	-16
TOTAL RESPONSE TENDENCIES	-15	-19	-10	-19	-15

Fig. 5. Response tendencies on third set of trials

TABLE 1
DISTRIBUTION OF CHIMPANZEES' RESPONSES IN A SEVEN-CHOICE APPARATUS
[from Spence (9)]

Subjects	Boxes						
	1	2	3	4	5	6	7
Bentia			1	14	5		
Soda			5	13	2		
Mamo			4	12	4		
Wendy			11	7	2		
Nira			5	9	6		
Al			5	6	8	1	
Josie		2	7	8	3		
Cuba			9	10	1		
Mona		2	7	8	3		
Pati		4	6	6	3	1	
Total		8	60	93	37	2	

certain inhibitory tendencies. But one response that occurs on every set of trials is the correct response. Hence the excitatory tendencies to make the correct response gradually gain strength at the expense of the inhibitory tendencies that operate to block the correct response. As the theory predicts, where the correct response alternates from side to side, learning will occur very slowly.

While this theory permits of rather pat predictions in the cases of the two phenomena that have been examined, the question of how Sadovinkova's siskin (11) was able to solve a problem that baffles many chimpanzees is one that cannot be answered conclusively without knowing exactly what responses the bird made and in what order it made them. In the light of the foregoing discussion it might be suggested that an answer may be found by applying a gradient theory to the raw data for the individual subject. By putting together a gradient theory and the fact that a subject does not always make all the incorrect responses before making the correct one, it may be possible to explain quite easily all those puzzling phylogenetic reversals such as this one, in which a bird can solve a problem that many a chimpanzee can not.

Spence (9), however, was able to get chimpanzees to solve the middle-door problem. They did not, of course, go to the middle door 100 per cent of the time; sometimes they went to other doors. But the sample of the data on their responses that Spence gives seems to lend support to a gradient theory. It is reproduced in Table 1. Spence used a seven-choice apparatus in which the two end doors were blocked off, giving him a five-door setting for the problem.

Apropos of these results Spence remarks:

> One very striking fact revealed in this table is the extent to which all of the subjects tended to respond to one of the three middle boxes of the settings, as though they had "aimed" their response at the middle or center box but lacked precision in execution (9, p. 24).

What strikes the present writer about the data is that they are in accord with the gradient theory. To be sure, the ratio between 93 and 37 is not quite 3 to 1, but almost. And the 8 and the 2 seem to argue in favor of a wider gradient than the one assumed in this paper. The only deviation that might militate seriously against the idea of a gradient is the lack of symmetry in the curve. The animals went to No. 3 60 times, as against 37 times for No. 5. But this tendency to reach to the left may be due to nothing more than their being repeatedly hand-fed by right-handed psychologists.

The third phenomenon upon which this gradient hypothesis may shed some light is the manner in which Spence's chimpanzees changed from one response to another. After being taught to respond to the second from the left end of six doors, the animals were switched to a problem in which the correct door was

the right-end door of seven doors, and it was noted that the animals shifted over gradually, trying each door in succession from left to right, although occasionally an animal would skip a door. Now, while the theory has no ready explanation for the skipping, it would seem in this process of gradual shifting that the animals were moving down a gradient, building up inhibition with each successive nonreinforcement and thus never having a tendency to back up again. Let us say that the animal keeps responding to No. 2, which had been the correct door, until the inhibitory potential outweighs the excitatory. Then the animal shifts to No. 3 which still has some excitatory potential left which generalized from No. 2 when No. 2 was the correct door. Nonreinforcement at No. 3 piles up inhibition not only at No. 3 itself but also at No. 2 and No. 4 in equal quantities. But inhibition at No. 2 was greater than at No. 3, and much greater than at No. 4, before the shift, so now No. 4 has the greatest tendency to evoke a response and the animal shifts to No. 4. This process is repeated until the chimpanzee finally reaches the end door, which is the correct one, and receives some positive reinforcement.

The fourth phenomenon to be explained is the tendency of the animals to overshoot the mark in shifting toward the correct door when the correct door is the second one from the end. All of Spence's subjects exhibited this tendency. If an animal is shifting from left to right and the correct door is No. 4, next to the end door, No. 5, the chimpanzee reaches No. 4, is rewarded, but on the next try goes to No. 5 instead of staying at No. 4. Oddly enough, this is exactly what the theory would predict. Let us suppose that in the process of shifting from left to right the animal has reached No. 3. In addition to the inhibitory potential it acquired from the nonreinforced response to No. 2, No. 3 now has an inhibitory potential in its own right of, let us say, − 10. This generalizes to No. 4 to the extent of, say, − 6, and the gradient falls to zero at No. 5. The animal now makes a response to No. 4 and is reinforced. This gives No. 4 an excitatory potential of + 15 and it generalizes to No. 3 and No. 5 to the extent of + 11. The response tendencies after the reinforced response to No. 4 are thus:

	No. 3	No. 4	No. 5
No. 2 gradient	− 6	0	0
No. 3 gradient	−10	− 6	0
No. 4 gradient(+)	+11	+15	+11
Response tendencies	− 5	+ 9	+11

Hence it can be predicted in spite of the reinforcement at No. 4 that the next response will be to door No. 5. After a nonreinforced response to No. 5, door No. 4 will have the edge, and the animal will then make the correct response, for No. 3 will then have a resultant value of −5, the same as before; No. 4 will have + 3 and No. 5 will be reduced to + 1. As far as this phenomenon of overshooting the mark is concerned, the data seem to be adequately accounted for by the hypothesis.

While this gradient theory appears to have some explanatory value, and this paper has attempted to show that it has, it is not intended that this hypothesis be contradictory to all other theories of multiple-choice learning. It does, however, take the problem out of the realm of such concepts as ideation and link it to such other sorts of learning as conditioning and discrimination learning, and that much was intended. But this is not to say that this gradient theory is a complete theory of multiple-choice learning. Presumably, multiple-choice learning, like any other kind of

learning, is subject to such factors as drive strength, delay of reinforcement, amount of reward, etc. (3), and a more adequate theory than the present one will be a theory that synthesizes all of these things.

In summary, then, this gradient theory of multiple-choice learning assumes that responses are the outcome of excitation due to reinforcement and inhibition due to nonreinforcement. Both excitation and inhibition generalize in a gradient-like fashion, and they interact in such a way that the strength of the response tendency is the algebraic sum of excitation (positive) and inhibition (negative). On the basis of this theory, explanations are offered for four phenomena of multiple-choice learning: (*a*) the fact that animals find the middle-door problem harder than the end-door problem, (*b*) the fact that human subjects find it very difficult when the correct choice alternates about the middle, (*c*) the fact that in chimpanzees the shift from a formerly correct response to a new correct response is a gradual approach process, and (*d*) the fact that in so shifting, chimpanzees have a tendency to overshoot the mark and respond to the next door just after being rewarded at the correct door.

REFERENCES

1. BURTT, H. E. A study of the behavior of the white rat by the multiple-choice method. *J. anim. Behav.*, 1916, **6**, 222–246.
2. HAMILTON, G. V. A study of trial-and-error reactions in mammals. *J. anim. Behav.*, 1911, **1**, 33–66.
3. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. HUNTER, W. S. Yerkes' human and animal behavior. *Psychol. Bull.*, 1916, **13**, 327–330.
5. MULLER, D. Sinnesphysiologische und psychologische Untersuchungen an Musteliden. *Z. vergl. Physiol.*, 1930, **12**, 293–328.
6. MUNN, N. L. *Psychology.* New York: Houghton-Mifflin, 1946.
7. SADOVINKOVA, M. P. A study of the behavior of birds by the multiple choice method. *J. comp. Psychol.*, 1923, **3**, 249–282.
8. SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
9. SPENCE, K. W. The solution of multiple-choice problems by chimpanzees. *Comp. Psychol. Monogr.*, 1939, **15**, No. 3.
10. TELLIER, M. Le macaque sensit-il le rapport logique? *Bull. Soc. Sci., Liége,* 1932, No. 2.
11. UNDERWOOD, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.
12. WASHBURN, M. F. *The animal mind.* New York: Macmillan, 1936.
13. YERKES, R. M. The mental life of monkeys and apes: a study of ideational behavior. *Behav. Monogr.*, 1916, **3**, 1–145.
14. YERKES, R. M. Methods of exhibiting reactive tendencies characteristic of ontogenetic and phylogenetic stages. *J. anim. Behav.*, 1917, **7**, 11–28.
15. YERKES, R. M. Modes of behavioral adaptation in chimpanzee to multiple-choice problems. *Comp. Psychol. Monogr.*, 1934, **10**, No. 1.
16. YERKES, R. M., & COBURN, C. A. A study of the pig *sus scrofa* by the multiple choice method. *J. anim. Behav.*, 1915, **5**, 185–225.

[MS. received October 30, 1951]

Psychological Review
Vol. 60, No. 1, 1953

MICE, MONKEYS, MEN, AND MOTIVES [1]

HARRY F. HARLOW

University of Wisconsin

Many of psychology's theoretical growing pains—or, in modern terminology, conditioned anxieties—stem from the behavioral revolution of Watson. The new psychology intuitively disposed of instincts and painlessly disposed of hedonism. But having completed this St. Bartholomew–type massacre, behavioristic motivation theory was left with an aching void, a nonhedonistic aching void, needless to say.

Before the advent of the Watsonian scourge the importance of external stimuli as motivating forces was well recognized. Psychologists will always remain indebted to Loeb's (21) brilliant formulation of tropistic theory, which emphasized, and probably overemphasized, the powerful role of external stimulation as the primary motivating agency in animal behavior. Unfortunately, Loeb's premature efforts to reduce all behavior to overly simple mathematical formulation, his continuous acceptance of new tropistic constructs in an effort to account for any aberrant behavior not easily integrated into his original system, and his abortive attempt to encompass all behavior into a miniature theoretical system doubtless led many investigators to underestimate the value of his experimental contributions.

Thorndike (30) was simultaneously giving proper emphasis to the role of external stimulation as a motivating force in learning and learned performances. Regrettably, these motivating processes were defined in terms of pain and pleasure, and it is probably best for us to dispense with such lax, ill-defined, subjective terms as pain, pleasure, anxiety, frustration, and hypotheses—particularly in descriptive and theoretical rodentology.

Instinct theory, for all its terminological limitations, put proper emphasis on the motivating power of external stimuli; for, as so brillantly described by Watson (31) in 1941, the instinctive response was elicited by "serial stimulation," much of which was serial external stimulation.

The almost countless researches on tropisms and instincts might well have been expanded to form a solid and adequate motivational theory for psychology—a theory with a proper emphasis on the role of the external stimulus and an emphasis on the importance of incentives as opposed to internal drives per se.

It is somewhat difficult to understand how this vast and valuable literature was to become so completely obscured and how the importance of the external stimulus as a motivating agent was to become lost. Pain-pleasure theory was discarded because the terminology had subjective, philosophical implications. Instinct theory fell into disfavor because psychologists rejected the dichotomized heredity-environment controversy and, also, because the term "instinct" had more than one meaning. Why tropistic theory disappeared remains a mystery, particularly inasmuch as most of the researches were carried out on subprimate animal forms.

Modern motivation theory apparently evolved from an overpopularization of certain experimental and theoretical ma-

[1] This paper was presented September 3, 1951, as the presidential address of the Division of Experimental Psychology at the Chicago meetings of the American Psychological Association.

terials. Jennings' (14) demonstration that "physiological state" played a role in determining the behavior of the lower animal was given exaggerated importance and emphasis, thereby relegating the role of external stimulation to a secondary position as a force in motivation. The outstanding work in the area of motivation between 1920 and 1930 related to visceral drives and drive cycles and was popularized by Richter's idealized theoretical paper on "Animal Behavior and Internal Drives" (26) and Cannon's *The Wisdom of the Body* (3).

When the self-conscious behavior theorists of the early thirties looked for a motivation theory to integrate with their developing learning constructs, it was only natural that they should choose the available tissue-tension hypotheses. Enthusiastically and uncritically the S–R theorists swallowed these theses whole. For fifteen years they have tried to digest them, and it is now time that these theses be subjected to critical examination, analysis, and evaluation. We do not question that these theses have fertilized the field of learning, but we do question that the plants that have developed are those that will survive the test of time.

It is my belief that the theory which describes learning as dependent upon drive reduction is false, that internal drive as such is a variable of little importance to learning, and that this small importance steadily decreases as we ascend the phyletic scale and as we investigate learning problems of progressive complexity. Finally, it is my position that drive-reduction theory orients learning psychologists to attack problems of limited importance and to ignore the fields of research that might lead us in some foreseeable future time to evolve a theoretical psychology of learning that transcends any single species or order.

There can be no doubt that the single-celled organisms such as the amoeba and the paramecium are motivated to action both by external and internal stimuli. The motivation by external stimulation gives rise to heliotropisms, chemotropisms, and rheotropisms. The motivation by internal stimulation produces characteristic physiological states which have, in turn, been described as chemotropisms. From a phylogenetic point of view, moreover, neither type of motive appears to be more basic or more fundamental than the other. Both types are found in the simplest known animals and function in interactive, rather than in dominant-subordinate, roles.

Studies of fetal responses in animals from opossum to man give no evidence suggesting that the motivation of physiological states precedes that of external incentives. Tactual, thermal, and even auditory and visual stimuli elicit complex patterns of behavior in the fetal guinea pig, although this animal has a placental circulation which should guarantee against thirst or hunger (4). The newborn opossum climbs up the belly of the female and into the pouch, apparently in response to external cues; if visceral motives play any essential role, it is yet to be described (20). The human fetus responds to external tactual and nociceptive stimuli at a developmental period preceding demonstrated hunger or thirst motivation. Certainly, there is no experimental literature to indicate that internal drives are ontogenetically more basic than exteroceptive motivating agencies.

Tactual stimulation, particularly of the cheeks and lips, elicits mouth, head, and neck responses in the human neonate, and there are no data demonstrating that these responses are conditioned, or even dependent, upon physiological drive states. Hunger appears to lower the threshold for these responses to tactual stimuli. Indeed, the main role of

the primary drive seems to be one of altering the threshold for precurrent responses. Differentiated sucking response patterns have been demonstrated to quantitatively varied thermal and chemical stimuli in the infant only hours of age (15), and there is, again, no reason to believe that the differentiation could have resulted from antecedent tissue-tension reduction states. Taste and temperature sensations induced by the temperature and chemical composition of the liquids seem adequate to account for the responses.

There is neither phylogenetic nor ontogenetic evidence that drive states elicit more fundamental and basic response patterns than do external stimuli; nor is there basis for the belief that precurrent responses are more dependent upon consummatory responses than are consummatory responses dependent upon precurrent responses. There is no evidence that the differentiation of the innate precurrent responses is more greatly influenced by tissue-tension reduction than are the temporal ordering and intensity of consummatory responses influenced by conditions of external stimulation.

There are logical reasons why a drive-reduction theory of learning, a theory which emphasizes the role of internal, physiological-state motivation, is entirely untenable as a motivational theory of learning. The internal drives are cyclical and operate, certainly at any effective level of intensity, for only a brief fraction of any organism's waking life. The classical hunger drive physiologically defined ceases almost as soon as food—or nonfood—is ingested. This, as far as we know, is the only case in which a single swallow portends anything of importance. The temporal brevity of operation of the internal drive states obviously offers a minimal opportunity for conditioning and a maximal opportunity for extinction. The human being, at least in the continental United States, may go for days or even years without ever experiencing true hunger or thirst. If his complex conditioned responses were dependent upon primary drive reduction, one would expect him to regress rapidly to a state of tuitional oblivion. There are, of course, certain recurrent physiological drive states that are maintained in the adult. But the studies of Kinsey (17) indicate that in the case of one of these there is an inverse correlation between presumed drive strength and scope and breadth of learning, and in spite of the alleged reading habits of the American public, it is hard to believe that the other is our major source of intellectual support. Any assumption that derived drives or motives can account for learning in the absence of primary drive reduction puts an undue emphasis on the strength and permanence of derived drives, at least in subhuman animals. Experimental studies to date indicate that most derived drives (24) and second-order conditioned responses (25) rapidly extinguish when the rewards which theoretically reduce the primary drives are withheld. The additional hypothesis of functional autonomy of motives, which could bridge the gap, is yet to be demonstrated experimentally.

The condition of strong drive is inimical to all but very limited aspects of learning—the learning of ways to reduce the internal tension. The hungry child screams, closes his eyes, and is apparently oblivious to most of his environment. During this state he eliminates response to those aspects of his environment around which all his important learned behaviors will be based. The hungry child is a most incurious child, but after he has eaten and become thoroughly sated, his curiosity and all the learned responses associated with his curiosity take place. If this learning is conditioned to an internal drive state,

we must assume it is the resultant of backward conditioning. If we wish to hypothesize that backward conditioning is dominant over forward conditioning in the infant, it might be possible to reconcile fact with S–R theory. It would appear, however, that alternate theoretical possibilities should be explored before the infantile backward conditioning hypothesis is accepted.

Observations and experiments on monkeys convinced us that there was as much evidence to indicate that a strong drive state inhibits learning as to indicate that it facilitates learning. It was the speaker's feeling that monkeys learned most efficiently if they were given food before testing, and as a result, the speaker routinely fed his subjects before every training session. The rhesus monkey is equipped with enormous cheek pouches, and consequently many subjects would begin the educational process with a rich store of incentives crammed into the buccal cavity. When the monkey made a correct response, it would add a raisin to the buccal storehouse and swallow a little previously munched food. Following an incorrect response, the monkey would also swallow a little stored food. Thus, both correct and incorrect responses invariably resulted in S–R theory drive reduction. It is obvious that under these conditions the monkey cannot learn, but the present speaker developed an understandable skepticism of this hypothesis when the monkeys stubbornly persisted in learning, learning rapidly, and learning problems of great complexity. Because food was continuously available in the monkey's mouth, an explanation in terms of differential fractional anticipatory goal responses did not appear attractive. It would seem that the Lord was simply unaware of drive-reduction learning theory when he created, or permitted the gradual evolution of, the rhesus monkey.

The langurs are monkeys that belong to the only family of primates with sacculated stomachs. There would appear to be no mechanism better designed than the sacculated stomach to induce automatically prolonged delay of reinforcement defined in terms of homeostatic drive reduction. Langurs should, therefore, learn with great difficulty. But a team of Wisconsin students has discovered that the langurs in the San Diego Zoo learn at a high level of monkey efficiency. There is, of course, the alternative explanation that the inhibition of hunger contractions in multiple stomachs is more reinforcing than the inhibition of hunger contractions in one. Perhaps the quantification of the gastric variable will open up great new vistas of research.

Actually, the anatomical variable of diversity of alimentary mechanisms is essentially uncorrelated with learning to food incentives by monkeys and suggests that learning efficiency is far better related to tensions in the brain than in the belly.

Experimental test bears out the fact that learning performance by the monkey is unrelated to the theoretical intensity of the hunger drive. Meyer (23) tested rhesus monkeys on discrimination-learning problems under conditions of maintenance-food deprivation of 1.5, 18.5, and 22.5 hours and found no significant differences in learning or performance. Subsequently, he tested the same monkeys on discrimination-reversal learning following 1, 23, and 47 hours of maintenance-food deprivation and, again, found no significant differences in learning or in performance as measured by activity, direction of activity, or rate of responding. There was some evidence, not statistically significant, that the most famished subjects were a bit overeager and that in-

tense drive exerted a mildly inhibitory effect on learning efficiency.

Meyer's data are in complete accord with those presented by Birch (1), who tested six young chimpanzees after 2, 6, 12, 24, and 48 hr. of food deprivation and found no significant differences in proficiency of performance on six patterned string problems. Observational evidence led Birch to conclude that intense food deprivation adversely affected problem solution because it led the chimpanzee to concentrate on the goal to the relative exclusion of the other factors.

It may be stated unequivocally that, regardless of any relationship that may be found for other animals, there are no data indicating that intensity of drive state and the presumably correlated amount of drive reduction are positively related to learning efficiency in primates.

In point of fact there is no reason to believe that the rodentological data will prove to differ significantly from those of monkey, chimpanzee, and man. Strassburger (29) has recently demonstrated that differences in food deprivation from 5 hours to 47 hours do not differentially affect the habit strength of the bar-pressing response as measured by subsequent resistance to extinction. Recently, Sheffield and Roby (28) have demonstrated learning in rats in the absence of primary drive reduction. Hungry rats learned to choose a maze path leading to a saccharin solution, a nonnutritive substance, in preference to a path leading to water. No study could better illustrate the predominant role of the external incentive–type stimulus on the learning function. These data suggest that, following the example of the monkey, even the rats are abandoning the sinking ship of reinforcement theory.

The effect of intensity of drive state on learning doubtless varies as we ascend the phyletic scale and certainly varies, probably to the point of almost complete reversal, as we pass from simple to complex problems, a point emphasized some years ago in a theoretical article by Maslow (22). Intensity of nociceptive stimulation may be positively related to speed of formation of conditioned avoidance responses in the monkey, but the use of intense nociceptive stimulation prevents the monkey from solving any problem of moderate complexity. This fact is consistent with a principle that was formulated and demonstrated experimentally many years ago as the Yerkes-Dodson law (32). There is, of course, no reference to the Yerkes-Dodson law by any drive-reduction theorist.

We do not mean to imply that drive state and drive-state reduction are unrelated to learning; we wish merely to emphasize that they are relatively unimportant variables. Our primary quarrel with drive-reduction theory is that it tends to focus more and more attention on problems of less and less importance. A strong case can be made for the proposition that the importance of the psychological problems studied during the last fifteen years has decreased as a negatively accelerated function approaching an asymptote of complete indifference. Nothing better illustrates this point than the kinds of apparatus currently used in "learning" research. We have the single-unit T-maze, the straight runway, the double-compartment grill box, and the Skinner box. The single-unit T-maze is an ideal apparatus for studying the visual capacities of a nocturnal animal; the straight runway enables one to measure quantitatively the speed and rate of running from one dead end to another; the double-compartment grill box is without doubt the most efficient torture chamber which is still legal; and the Skinner box enables one to demonstrate discrimination learning in a greater number of trials than is required by any

other method. But the apparatus, though inefficient, give rise to data which can be splendidly quantified. The kinds of learning problems which can be efficiently measured in these apparatus represent a challenge only to the decorticate animal. It is a constant source of bewilderment to me that the neobehaviorists who so frequently belittle physiological psychology should choose apparatus which, in effect, experimentally decorticate their subjects.

The Skinner box is a splendid apparatus for demonstrating that the rate of performance of a learned response is positively related to the period of food deprivation. We have confirmed this for the monkey by studying rate of response on a modified Skinner box following 1, 23, and 47 hr. of food deprivation. Increasing length of food deprivation is clearly and positively related to increased rate of response. This functional relationship between drive states and responses does not hold, as we have already seen, for the monkey's behavior in discrimination learning or in acquisition of any more complex problem. The data, however, like rat data, are in complete accord with Crozier's (6) finding that the acuteness of the radial angle of tropistic movements in the slug Limax is positively related to intensity of the photic stimulation. We believe there is generalization in this finding, and we believe the generalization to be that the results from the investigation of simple behavior may be very informative about even simpler behavior but very seldom are they informative about behavior of greater complexity. I do not want to discourage anyone from the pursuit of the psychological Holy Grail by the use of the Skinner box, but as far as I am concerned, there will be no moaning of farewell when we have passed the pressing of the bar.

In the course of human events many psychologists have children, and these children always behave in accord with the theoretical position of their parents. For purposes of scientific objectivity the boys are always referred to as "Johnny" and the girls as "Mary." For some eleven months I have been observing the behavior of Mary X. Perhaps the most striking characteristic of this particular primate has been the power and persistence of her curiosity-investigatory motives. At an early age Mary X demonstrated a positive valence to parental thygmotatic stimulation. My original interpretation of these tactual-thermal erotic responses as indicating parental affection was dissolved by the discovery that when Mary X was held in any position depriving her of visual exploration of the environment, she screamed; when held in a position favorable to visual exploration of the important environment, which did not include the parent, she responded positively. With the parent and position held constant and visual exploration denied by snapping off the electric light, the positive responses changed to negative, and they returned to positive when the light was again restored. This behavior was observed in Mary X, who, like any good Watson child, showed no "innate fear of the dark."

The frustrations of Mary X appeared to be in large part the results of physical inability to achieve curiosity-investigatory goals. In her second month, frustrations resulted from inability to hold up her head indefinitely while lying prone in her crib or on a mat and the consequent loss of visual curiosity goals. Each time she had to lower her head to rest, she cried lustily. At nine weeks attempts to explore (and destroy) objects anterior resulted in wriggling backward away from the lure and elicited violent negative responses. Once she negotiated forward locomotion, exploration set in, in earnest, and, much

to her parents' frustration, shows no sign of diminishing.

Can anyone seriously believe that the insatiable curiosity-investigatory motivation of the child is a second-order or derived drive conditioned upon hunger or sex or any other internal drive? The S–R theorist and the Freudian psychoanalyst imply that such behaviors are based on primary drives. An informal survey of neobehaviorists who are also fathers (or mothers) reveals that all have observed the intensity and omnipresence of the curiosity-investigatory motive in their own children. None of them seriously believes that the behavior derives from a second-order drive. After describing their children's behavior, often with a surprising enthusiasm and frequently with the support of photographic records, they trudge off to their laboratories to study, under conditions of solitary confinement, the intellectual processes of rodents. Such attitudes, perfectly in keeping with drive-reduction theory, no doubt account for the fact that there are no experimental or even systematic observational studies of curiosity-investigatory–type external-incentive motives in children.

A key to the real learning theory of any animal species is knowledge of the nature and organization of the unlearned patterns of response. The differences in the intellectual capabilities of cockroach, rat, monkey, chimpanzee, and man are as much a function of the differences in the inherent patterns of response and the differences in the inherent motivational forces as they are a function of sheer learning power. The differences in these inherent patterns of response and in the motivational forces will, I am certain, prove to be differential responsiveness to external stimulus patterns. Furthermore, I am certain that the variables which are of true, as opposed to psychophilosophical, importance are not constant from learning problem to learning problem even for the same animal order, and they are vastly diverse as we pass from one animal order to another.

Convinced that the key to human learning is not the conditioned response but, rather, motivation aroused by external stimuli, the speaker has initiated researches on curiosity-manipulation behavior as related to learning in monkeys (7, 10, 12). The justification for the use of monkeys is that we have more monkeys than children. Furthermore, the field is so unexplored that a systematic investigation anywhere in the phyletic scale should prove of methodological value. The rhesus monkey is actually a very incurious and nonmanipulative animal compared with the anthropoid apes, which are, in turn, very incurious nonmanipulative animals compared with man. It is certainly more than coincidence that the strength and range of curiosity-manipulative motivation and position within the primate order are closely related.

We have presented three studies which demonstrate that monkeys can and do learn to solve mechanical puzzles when no motivation is provided other than presence of the puzzle. Furthermore, we have presented data to show that once mastered, the sequence of manipulations involved in solving these puzzles is carried out relatively flawlessly and extremely persistently. We have presented what we believe is incontrovertible evidence against a second-order drive interpretation of this learning.

A fourth study was carried out recently by Gately at the Wisconsin laboratories. Gately directly compared the behavior of two groups of four monkeys presented with banks of four identical mechanical puzzles, each utilizing three restraining devices. All four food- plus puzzle-rewarded monkeys solved the four identical puzzles, and only one of the four monkeys motivated by curiosity

alone solved all the puzzles. This one monkey, however, learned as rapidly and as efficiently as any of the food-rewarded monkeys. But I wish to stress an extremely important observation made by Gately and supported by quantitative records. When the food-rewarded monkeys had solved a puzzle, they abandoned it. When the nonfood-rewarded animals had solved the puzzle, they frequently continued their explorations and manipulations. Indeed, one reason for the nonfood-rewarded monkeys' failure to achieve the experimenter's concept of solution lay in the fact that the monkey became fixated in exploration and manipulation of limited puzzle or puzzle-device components. From this point of view, hunger-reduction incentives may be regarded as motivation-destroying, not motivation-supporting, agents.

Twenty years ago at the Vilas Park Zoo, in Madison, we observed an adult orangutan given two blocks of wood, one with a round hole, one with a square hole, and two plungers, one round and one square. Intellectual curiosity alone led it to work on these tasks, often for many minutes at a time, and to solve the problem of inserting the round plunger in both holes. The orangutan never solved the problem of inserting the square peg into the round hole, but inasmuch as it passed away with perforated ulcers a month after the problem was presented, we can honestly say that it died trying. And in defense of this orangutan, let it be stated that it died working on more complex problems than are investigated by most present-day learning theorists.

Schiller [2] has reported that chimpanzees solve multiple-box–stacking problems without benefit of food rewards, and he has presented observational evidence that the joining of sticks resulted from manipulative play responses.

[2] Personal communication.

The Cebus monkey has only one claim to intellectual fame—an ability to solve instrumental problems that rivals the much publicized ability of the anthropoid apes (11, 18). It can be no accident that the Cebus monkey, inferior to the rhesus on conventional learning tasks, demonstrates far more spontaneous instrumental-manipulative responses than any old-world form. The complex, innate external-stimulus motives are variables doubtlessly as important as, or more important than, tissue tensions, stimulus generalization, excitatory potential, or secondary reinforcement. It is the oscillation of sticks, not cortical neurons, that enables the Cebus monkey to solve instrumental problems.

No matter how important may be the analysis of the curiosity-manipulative drives and the learning which is associated with them, we recognize the vast and infinite technical difficulties that are inherent in the attack on the solution of these problems—indeed, it may be many years before we can routinely order such experiments in terms of latin squares and factorial designs, the apparent *sine qua non* for publication in the *Journal of Experimental Psychology* and the *Journal of Comparative and Physiological Psychology.*

There is, however, another vast and important area of external-stimulus incentives important to learning which has been explored only superficially and which can, and should, be immediately and systematically attacked by rodentologists and primatologists alike. This is the area of food incentives—or, more broadly, visuo-chemo variables—approached from the point of view of their function as motivating agents per se. This function, as the speaker sees it, is primarily an affective one and only secondarily one of tissue-tension reduction. To dispel any fear of subjectivity, let us state that the affective tone of food incentives can probably be scaled

by preference tests with an accuracy far exceeding any scaling of tissue tensions. Our illusion of the equal-step intervals of tissue tensions is the myth that length of the period of deprivation is precisely related to tissue-tension intensity, but the recent experiments by Koch and Daniel (19) and Horenstein (13) indicate that this is not true, thus beautifully confirming the physiological findings of thirty years ago.

Paired-comparison techniques with monkeys show beyond question that the primary incentive variables of both differential quantity and differential quality can be arranged on equal-step scales, and there is certainly no reason to believe that variation dependent upon subjects, time, or experience is greater than that dependent upon physiological hunger.

In defense of the rat and its protagonists, let it be stated that there are already many experiments on this lowly mammal which indicate that its curiosity-investigatory motives and responsiveness to incentive variables can be quantitatively measured and their significant relationship to learning demonstrated. The latent learning experiments of Buxton (2), Haney (9), Seward, Levy, and Handlon (27), and others have successfully utilized the exploratory drive of the rat. Keller (16) and Zeaman and House (35) have utilized the rat's inherent aversion to light, or negative heliotropistic tendencies, to induce learning. Flynn and Jerome (8) have shown that the rat's avoidance of light is an external-incentive motivation that may be utilized to obtain the solution of complex learned performances. For many rats it is a strong and very persistent form of motivation. The importance of incentive variables in rats has been emphasized and re-emphasized by Young (33), and the influence of incentive variables on rat learning has been demonstrated by Young (33), Zeaman (34), Crespi (5), and others. I am not for one moment disparaging the value of the rat as a subject for psychological investigation; there is very little wrong with the rat that cannot be overcome by the education of the experimenters.

It may be argued that if we accept the theses of this paper, we shall be returning to an outmoded psychology of tropisms, instincts, and hedonism. There is a great deal of truth to this charge. Such an approach might be a regression were it not for the fact that psychology now has adequate techniques of methodology and analysis to attack quantifiably these important and neglected areas. If we are ever to have a comprehensive theoretical psychology, we must attack the problems whose solution offers hope of insight into human behavior, and it is my belief that if we face our problems honestly and without regard to, or fear of, difficulty, the theoretical psychology of the future will catch up with, and eventually even surpass, common sense.

REFERENCES

1. Birch, H. C. The relation of previous experience to insightful problem solving. *J. comp. Psychol.*, 1945, 39, 15–22.
2. Buxton, C. E. Latent learning and the goal gradient hypothesis. *Contr. psychol. Theor.*, 1940, 2, No. 2.
3. Cannon, W. B. *The wisdom of the body.* New York: Norton, 1932.
4. Carmichael, L. An experimental study in the prenatal guinea-pig of the origin and development of reflexes and patterns of behavior in relation to the stimulation of specific receptor areas during the period of active fetal life. *Genet. Psychol. Monogr.*, 1934, 16, 337–491.
5. Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, 55, 467–517.
6. Crozier, W. J. The study of living organisms. In C. Murchison (Ed.), *The foundations of experimental psychology.* Worcester, Mass.: Clark Univer. Press, 1929.

7. DAVIS, R. T., SETTLAGE, P. H., & HARLOW, H. F. Performance of normal and brain-operated moneys on mechanical puzzles with and without food incentive. *J. genet. Psychol.*, 1950, 77, 305–311.
8. FLYNN, J. P., & JEROME, E. A. Learning in an automatic multiple-choice box with light as incentive. *J. comp. physiol. Psychol.*, 1952, 45, 336–340.
9. HANEY, G. W. The effect of familiarity on maze performance of albino rats. *Univer. Calif. Publ. Psychol.*, 1931, 4, 319–333.
10. HARLOW, H. F. Learning and satiation of response in intrinsically motivated complex puzzle performance by monkeys. *J. comp. physiol. Psychol.*, 1950, 43, 289–294.
11. HARLOW, H. F. Primate learning. In C. P. Stone (Ed.), *Comparative psychology.* (3rd Ed.) New York: Prentice-Hall, 1951.
12. HARLOW, H. F., HARLOW, MARGARET K., & MEYER, D. R. Learning motivated by a manipulation drive. *J. exp. Psychol.*, 1950, 40, 228–234.
13. HORENSTEIN, BETTY. Performance of conditioned responses as a function of strength of hunger drive. *J. comp. physiol. Psychol.*, 1951, 44, 210–224.
14. JENNINGS, H. S. *Behavior of the lower organisms.* New York: Columbia Univer. Press, 1906.
15. JENSEN, K. Differential reactions to taste and temperature stimuli in newborn infants. *Genet. Psychol. Monogr.*, 1932, 12, 361–476.
16. KELLER, F. S. Light-aversion in the white rat. *Psychol. Rec.*, 1941, 4, 235–250.
17. KINSEY, A. C., POMEROY, W. B., & MARTIN, C. E. *Sexual behavior in the human male.* Philadelphia: W. B. Saunders, 1948.
18. KLÜVER, H. *Behavior mechanisms in monkeys.* Chicago: Univer. Chicago Press, 1933.
19. KOCH, S., & DANIEL, W. J. The effect of satiation on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*, 1945, 35, 162–185.
20. LANGWORTHY, O. R. The behavior of pouch-young opossums correlated with the myelinization of tracts in the nervous system. *J. comp. Neurol.*, 1928, 46, 201–248.
21. LOEB, J. *Forced movements, tropisms and animal conduct.* Philadelphia: Lippincott, 1918.
22. MASLOW, A. H. A theory of human motivation. *Psychol. Rev.*, 1943, 50, 370–396.
23. MEYER, D. R. Food deprivation and discrimination reversal learning by monkeys. *J. exp. Psychol.*, 1951, 41, 10–16.
24. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951.
25. PAVLOV, I. P. *Conditioned reflexes* (translated by G. V. Anrep). London: Oxford Univer. Press, 1927.
26. RICHTER, C. P. Animal behavior and internal drives. *Quart. Rev. Biol.*, 1927, 2, 307–343.
27. SEWARD, J. P., LEVY, N., & HANDLON, J. H., JR. Incidental learning in the rat. *J. comp. physiol. Psychol.*, 1950, 43, 240–251.
28. SHEFFIELD, F. D., & ROBY, T. B. Reward value of a non-nutrient sweet taste. *J. comp. physiol. Psychol.*, 1950, 43, 471–481.
29. STRASSBURGER, R. C. Resistance to extinction of a conditioned operant as related to drive level at reinforcement. *J. exp. Psychol.*, 1950, 40, 473–487.
30. THORNDIKE, E. L. *Animal intelligence.* New York: Macmillan, 1911.
31. WATSON, J. B. *Behavior: An introduction to comparative psychology.* New York: Holt, 1914.
32. YERKES, R. M., & DODSON, J. D. The relation of strength of stimulus to rapidity of habit formation. *J. comp. Neurol. Psychol.*, 1908, 18, 459–482.
33. YOUNG, P. T. Food-seeking drive, affective process, and learning. *Psychol. Rev.*, 1949, 56, 98–121.
34. ZEAMAN, D. Response latency as a function of amount of reinforcement. *J. exp. Psychol.*, 1949, 39, 466–483.
35. ZEAMAN, D., & HOUSE, BETTY J. Response latency at zero drive after varying numbers of reinforcements. *J. exp. Psychol.*, 1950, 40, 570–583.

[MS. received November 23, 1951]

Psychological Review
Vol. 60, No. 1, 1953

THE PLACE OF THEORY IN SCIENCE [1]

KARL M. DALLENBACH

University of Texas

I was led to select "The Place of Theory in Science" as the subject of my address this evening by observations made last year as I listened with you to our divisional symposium upon "Theories of Learning." On that occasion every speaker ardently defended a theory which he had himself espoused, marshalling evidence favorable to his own views, and attacking the proposals of his opponents. It was a lively and heated discussion, but its fervor was worthy of a better cause.

Consider some of the theories "ardently defended" in our youth: those pertaining to the "mind-body" relationship; those dealing with "play," where Herbert Spencer, Karl Gross, John Ruskin, and G. Stanley Hall were now extolled and again derided; and those concerned with the origin of language.

The theories of language naïvely discussed included the onomatopoeic or "bow-wow theory" which attributes the origin of language to the attempt of our primitive ancestors to imitate sounds heard and to use those imitations to represent the objects of their origin; the interjectional theory, called in derision by its opponents the "pooh-pooh theory," which supposes that language arose from the automatic exclamations or ejaculations of primitive man; the reflex theory, nicknamed the "ding-dong theory," which holds that various objects and situations in man's environment arouse unlearned or reflex gestures and vocalizations which, under repetition, become associated with them and ultimately to represent them; the social or "yo-he-ho theory" which maintains that language developed from vocalizations made by primitive men in the course of cooperative activities; and finally the theory which regards language as a form of expressive movement that accompanied and arose from gesture.

We were, of course, also conversant with all the classical and the newer theories of sensation, perception, feeling, emotion, attention, and the like, which have come down to the present. I mentioned in particular the mind-body theories and those of play and language because few students today are concerned with them and many have not even heard of them. They were, however, argued and championed in our youth as ardently as any theory today —even learning theory. The terse and uncomplimentary nicknames given the various theories of language impart, for example, the intensity of the feelings of the participants in that controversy.

What has become of those early youthful debates? Did criticism prove a leaven to bring about unanimity of opinion upon theory? Not a bit of it. The mind-body problem was not solved; it merely passed from the psychological scene. It was given over to philosophy and metaphysics and the doctrine of psychophysical parallelism was tacitly assumed—at least by many experimental psychologists, who, in order to get on with their proper work, resolved the problem by placing parallelism among their postulates. Like a fortified atoll, which has long resisted frontal attack, the mind-body problem was by-passed and when that was done, it developed that it need not be taken. Similarly, the origins of play and of language are no longer of concern. Their existence is self-evident and their study need not

[1] Address of the retiring president of the Division of General Psychology, American Psychological Association, Chicago, Illinois, August 31, 1951.

be delayed, as it was eventually discovered, for a verdict regarding their origin —which is still, after nearly fifty years, to be returned. The theories of black and of the elementariness of the intermediate colors, also long and vigorously debated, have similarly disappeared. Indeed, though I have searched long through the history of our science for a theory that gave way to criticism, I have been unable to find one.

Theories do not succumb to abstract argumentation. They do not die but—like an occasional old soldier—they just fade away. They pass from the scientific stage not because they have been discredited but because they have been superseded or by-passed—pushed off the stage and replaced by other theories.

Why then become enamored with theory and argue and debate it emotionally as if its issues were of vital concern? The chief and, as I believe, the only justifiable reasons are pride and the obligations of authorship. The author of a theory owes allegiance to his brain-child as to the child of his body. He is responsible for its elucidation and clarification, hence may be pardoned if he is emotional in its defense and now and then sarcastic and caustic in his reply—even as he would be in case his children were unjustly criticized (as he must believe) by a disagreeable (as he must think) neighbor. If he were wise, he might concede that his theory and even his children need correction. Some authors and parents are of sufficient stature to grant that and they modify their theories and the behavior of their children in the light of criticism received, but their number is few. Most authors defend their theories to the last jot and tittle; and most parents admit of no flaws, at least not in the face of criticism, among their "little angels."

If authors of theories were the only ones to become emotional in defense, the reasons for their fervor would not be far to seek, but others, who espouse and defend theory, are often as ardent and as zealous as the authors themselves. This may be due, as I have suggested, to youthful loyalties, or to the desire to hold fast to something in this rapidly changing and fluctuating world, even if it be merely theory. In either case, however, it is a mistake because there is such a thing as theory-blindness, i.e., being so blinded by adherence to a specific theory as to be unable truly to see the facts. When one defends theory, particularly if one is emotional about it, one is very apt to select the data which support it and be blind to all others that do not. Nowhere is this better shown than in the conflict between the Helmholtz and the Hering theories of vision. The results of experiments in vision, particularly upon color-blindness, color-mixing, and color zones, are determined by the theoretical prejudices of the investigators.

Owing to the training that students of my generation received, my loyalties among the various contending theories were early established. Fortunately for me, however, the exalted opinion that I had of theory in general and of my favorite theories in particular did not last for long. At my first conference as a graduate student with Titchener, I had the temerity to ask him in which theory of vision he believed. He replied, "Believe, believe, why, I don't believe in any." Then followed a discourse upon theory and its place in experiment which ended with the admonition, "Carry your theories lightly," by which he meant do not rush to their attack or to their defense but go about your own proper work of gathering facts. They and they alone verify or disprove theory.

This advice does not mean, of course, that theory should be avoided and not discussed nor criticized. Quite to the contrary, it means that every one within

his own special field of knowledge should know all the theories, love a few, wed none. The best way to learn to know a theory is to discuss it but keep the discussion centered upon fact. Keep your theories fluid; do not permit them to crystallize until they have been factually demonstrated or disproved.

Yes, carry theory lightly that you may follow where fact may lead; and that you may forsake a theory without suffering feelings of guilt. Free from those paralyzing inhibitions, you may discuss, criticize, and judge theory impartially, which is something you will never be able to do as long as you carry the defense of theory, like Sinbad's Old Man of the Sea, upon your shoulders. Theory takes over, if given the opportunity, and you may not be so fortunate as Sinbad in freeing yourself from that incubus once you have burdened yourself with it.

Controversy over theory is, moreover, futile. It advertises and calls attention to the theory criticized and it is not effective against the will to believe. It is useless, bootless, and to no purpose. History is replete with instances in which that is shown to be true.

Sir William Hamilton, for example, spent many of the best years of his life performing crucial experiments and writing polemical articles against phrenology. Did his studies and articles do any good? Did they hasten the demise of phrenology? Did they wean any one away from those false doctrines? Was Hamilton's time and labor spent to the best advantage? The answer to these questions must be "no." Phrenology endured as a controversial doctrine nearly forty years after Hamilton had written his last critical paper. It was the growth of knowledge; the results of the experimental work of the brain physiologists, Broca, Fritsch, Hitzig, Ferrier, Goltz, Munk, that assigned the doctrine of phrenology to yesterday's seven thousand years.

Again, for nearly fifty years after the appearance of the *Origin of Species* (1859) the biologists and churchmen were engaged in a bitter controversy over Darwin's theories and the dogmas of revealed religion. The time and effort spent in those debates was wasted. The bishops could not convince the biologists and the biologists could not convince the bishops. The debates were destined to be futile from the start. The bishops were loyal to their dogma and the facts and arguments of the biologists fell upon deaf ears; and the scientists were loyal to Darwin's theory and the dogma of the churchmen left them unmoved. If the biologists had carried their theories lightly, had ignored the churchmen's criticisms, had gone about their own proper business of experimenting and seeking facts, their science would not have been relatively sterile experimentally during the forty-year period following the appearance of Darwin's book. Mendel did just that. He carried on within the precincts of a monastery and gave us results of first magnitude which were overlooked and neglected in the intensity of the controversy being waged over theory. It was not until 1900, when Mendel's experiments and results were discovered and publicized by De Vries, that the biologists again went about their own proper work.

Criticism of Freudianism is, similarly, labor lost. We have had a tremendous amount of it during the last forty years but it has been futile. It is not read by those wedded to psychoanalytic dogma and it is not needed by those who are not.

Thus far in my discussion I have been using the term "theory" in its broad, general sense to denote any tentative explanation of phenomena. In addition to that meaning, "theory" has a restricted use. It is frequently limited to propositions that have received a considerable amount of verification, that have passed beyond the tentative stage. When that is the case, its broad, general meaning is given over to such terms as *conjecture, speculation, doctrine, hypothesis*.

Attempts have been made to define all of these terms rigidly and to arrange them in

a progressive series according to the increasing adequacy of the evidence for them. Beginning with *conjecture* (an inference drawn wholly from presumptive evidence), and going to *speculation* (a generalization suggested by available but insecure observations), to *doctrine* (a thesis advanced upon the basis of authority), to *hypothesis* (a preliminary assumption or principle adopted for the explanation of observed facts), to *theory* (a specific formula propounded on the basis of secure facts for the purpose of explaining a group of given phenomena), and on to *law* (a statement of the fundamental uniformity of the facts of observation), the evidence for each concept becomes more and more factual and secure. As desirable as the rigid definition and serial ordering of these terms may be, the attempts made to achieve them have not been accepted. These terms, particularly hypothesis and theory, are used loosely and indiscriminately. What one author calls a "theory" another calls an "hypothesis" and frequently you find indecision and an author writing of "theory or hypothesis."

Among these various terms and definitions, two differences appear that are aptly illustrated by the figures of speech used to denote their function. One is that they are "bridges," spanning gaps in our knowledge, which are constructed to bring together large groups of facts or phenomena. The other is that they are "tools" devised for and used on specific occasions, being restricted to the explanation of a small number or groups of facts.

The first, the "bridge" analogy, is amply illustrated by such examples as the nebular hypothesis, the doctrine of associationism, the doctrine of specific nerve energies, the doctrine of creative synthesis, the phi-gamma hypothesis. Though bridges may be small, like culverts under a road, we usually think of them as grand, majestic spans over vast tracts. For such conceptual structures, I am, following Newton's lead, using "hypothesis." Newton restricted this term to the larger explanations, because etymologically "hypothesis" means "foundation."

The second, the "tool" analogy, is illustrated by reference to any well-reported experiment. It is the specific proposition the experimenter had in mind when he designed his investigation. It may be large, like a crane or derrick, or small, like a tack hammer or needle. In either case, it has a specific purpose, i.e., the formulation or tentative statement of the problem which the study was undertaken to test. For such concepts I am using the term "theory."

Since bridges and tools may be large or small, there may be times when they overlap in size, when the analogies do not hold—a crane, a large tool, may, for example, be larger than a culvert, a small bridge. By and large, however, a bridge exceeds a tool in size and extent, hence the distinction, since we are dealing only with analogies, need not be strained.

As tools are used in the construction, maintenance, and continuation of bridges, so theories are deducible from and subsumed by hypotheses. It is always with reference to some larger hypothesis that theories are formulated and employed.

Hypotheses, in the "bridge" sense, though regarded lightly by every one, are of great importance to science. They not only guide and give sanction to theory and experiment but they also consolidate the factual gains as they are made by giving them a basis of classification. The last, a basis of classification, is by no means the least of their contributions though it is frequently overlooked and ignored by writers upon this topic. The Müllerian doctrine of the specific energy of nerves is a case in point. As I have once before pointed out (4), through the ages since Aristotle, various and numerous investigators have sought, under the traditional doctrine of the five senses, to analyze the fifth sense, the sense of touch, into several senses. Their efforts, in many instances surprisingly accurate, were, however, without lasting significance. Their achievements made little impression upon their own age and generation and were unknown to succeeding generations. Every au-

thor wrote, against the background of the traditional doctrine of the five senses, as if he were proposing something new. Lapses of centuries and differences in language do not account for the lack of continuity for Bell, Erasmus Darwin's contemporary in England, gave no indication of knowing that Darwin had proposed a "muscle sense" when he wrote of it. What was lacking all along to consolidate the gains as they were being made through the centuries was the positive principle of an hypothesis. Müller, though he himself held to the traditional doctrine of the five senses, gave it to us in his principle of the dependence of qualitative difference upon the specific energy of nerves.

We should of course prefer to work with hypotheses that turn out to be true, but any hypothesis, true or false, that is capable of being tested by experiment is better than none. If false, it may lead to the formulation of a better, which in turn may lead to one still better, and thus, by successive demolitions and reconstructions, an accurate hypothesis—true knowledge—is approached.

The recognition of the importance of hypotheses has led to attempts to safeguard them, i.e., to draw up rigid rules regarding the legitimacy of their formulation. Thus, an hypothesis is legitimate scientifically only when the phenomena to be explained actually exist. This precept, that you must first establish your fact before you conjecture its causes, may seem elemental, but history yields numerous instances of its violation. Poor, lonely, wretched, and doubtlessly insane, old women were tried and executed in this country not so many years ago upon the assumption, that is, upon the hypothesis, that they were witches. We need not be smug and feel that our times are superior for some of us today speculate about modern "witches" which, though dressed in new garb, have no greater existence in fact than the old. For example, instead of telepathy we have ESP; instead of a wide variety of faculties or powers we have censor, ego, superego, and id.[4]

I could continue, as can you, with a host of examples of the need for this rule, but I shall let one more, an historical and an amusing example, suffice.

Charles II of England, soon after the Royal Society was established under his patronage, sent it a request to explain the following phenomenon. "When a live fish is thrown into a basin of water, the basin, water, and fish do not weigh more than the basin and water before the fish is thrown in; whereas, when a dead fish is employed, the weight of the whole is exactly equal to the added weights of the basin, the water, and the fish" (5). To answer their king and patron, the members proceeded to study the problem and much learned discussion ensued. Elaborate papers, propounding various hypotheses, were read in explanation. At length a member, not satisfied with any of the hypotheses proposed, suggested that it might be expedient to see whether the facts were as the king had stated them. For this he was roundly condemned. Some members asserted that the "facts" were of such common knowledge that their test was not needed, and others declared that to doubt the "facts" was an insult to his majesty and to perform the experiment would be tantamount to an act of treason. Despite these objections, the experiment was made—when lo! to the confusion of the wise men of Gotham—the name by which the Society was then popularly known—it was found that the weight of the three, the basin, water, and fish, was identical whether a dead or a live fish was used.

A second rule regarding the legitimacy of an hypothesis is that it must not be barren, i.e., incapable of being subjected to the test of experiment, of being proved or disproved. Such hypotheses are of no use to science as they do not advance knowledge one iota. It now

and then turns out, however, that what is barren at one day and generation is not at the next when suitable instruments have been invented and adequate procedures devised. Thus, for example, most of the early hypotheses relating to air were barren and remained so until the air-pump was invented and the chemists had devised suitable methods of analysis. Since an hypothesis barren today may be fruitful tomorrow, much of the force of this second rule is lost.

Other rules governing the legitimacy of hypotheses are: they should not be at variance with known facts; they should not subvert that for which they were devised to explain; they should deal with real and not with hypothetical facts; they should not transcend experience; they should work simply and naturally and not involve the occult or supernatural in their explanations.

Though all of these rules are sound in the abstract, they have little practical value, in part because they are not widely known nor widely accepted, and in part because of the general belief that any rule which discourages hypothesis is undesirable.

Because we have a plethora of hypotheses in psychology, I sometimes sigh and wish that the rules governing their legitimacy were more widely known and strictly enforced. When I recall, however, what has happened in countries where a censorship has been applied to the hypotheses of science, I am content to let free enterprise reign. Hypotheses will never become a burden if we carry them lightly. We have the cure for the "disease" in our own hands.

Theory in the narrow, the "tool" sense, is the immediate basis of experimentation. It is the specific formulation of the problem—the overt reason for undertaking the work. By means of it the experiment is designed.

I was once charged and roundly criticized for undertaking an experiment "to support a theory" (6). Except for the verb "support," for which I substituted "test," I admitted the charge as I did not then and do not now see how an experimental problem can even be formulated, to say nothing of conducted and completed, without some sort of preceding theory (3).

Galilei's apocryphal experiment upon the velocity of falling bodies furnishes an apt illustration. According to report it was made to test the Aristotelian doctrine, which was widely accepted for many centuries, that heavy bodies fell faster than light bodies.

Galilei, as you will recall, is reported to have dropped simultaneously three bodies of widely varying weights from the top of the Leaning Tower of Pisa and to have found that they reached the ground, not in order of their weights, the heaviest first and the lightest last, but at the same time. We can only guess at what the theory was that prompted this experiment for neither he nor any of his contemporaries either reported or described this experiment (2).

Whatever our guess may be, the paradigm of a scientific experiment is that theory, under an inclusive hypothesis, comes before the mechanical work of collecting and treating the data. An experiment in science is the integration of these aspects. Unfortunately that is not always the case in psychology, for as Bentley has pointed out in his article on "The nature and uses of experiment in psychology" (1), the theoretical setting, the planning and arranging an event and its conditions, ranks high among the plural meanings attached today to an experiment within psychology. Bentley distinguishes no less than eight uses, the first five of which are without the benefit or sanction of a theory. Doing something, without a theory, is not

a scientific experiment. It is mere busy work. It contributes nothing to knowledge; it has no scientific purpose nor meaning. We thus escape the necessity, when such a report appears in print to clutter up our scientific literature, of taking it into account for, "if there is no meaning," as the king said to Alice in Wonderland, "that saves a world of trouble, you know, as we needn't try to find any."

REFERENCES

1. BENTLEY, M. The nature and uses of experiment in psychology. *Amer. J. Psychol.*, 1937, 50, 452–469.

2. COOPER, L. *Aristotle, Galileo and the Tower of Pisa.* Cornell Press: 1935, 1–102; Galileo and scientific history, *Sci. Mo.*, 1936, 43, 163–167.

3. DALLENBACH, K. M. Dr. Morgan on the measurement of attention. *Amer. J. Psychol.*, 1918, 29, 122–123.

4. DALLENBACH, K. M. Pain: history and present status. *Amer. J. Psychol.*, 1939, 52, 334 f.

5. HAMILTON, W. Lectures on metaphysics. Edinburgh and London: William Blackwood & Sons, 1870.

6. MORGAN, J. J. B. The overcoming of distraction and other resistances. *Arch. Psychol.*, 1916, 5, No. 35, 1–84.

[MS. received October 16, 1951]

Psychological Review
Vol. 60, No. 1, 1953

FACTS, PHENOMENA, AND FRAMES OF REFERENCE IN PSYCHOLOGY [1]

LOUIS O. KATTSOFF

University of North Carolina

A recent text book in psychology affirms boldly that facts, which are *true* in one frame of reference are *false* in another. "A fact, we find, is not an independent thing that we can memorize and depend upon and know that it will always be true. It is true only in its frame of reference, which means that it is false in others" (6, p. 4). This assertion is based on an idea now widely accepted and one which I have myself discussed in earlier papers, namely, that a fact is such only in a frame of reference (3, 4). In this paper I wish to examine the meaning of the statement "a fact is such only in a frame of reference" in order to bring out some ambiguities and clarify what I think is a basic misconception in Snygg and Combs. This task will also involve a discussion of frames of reference as well as the problem of the criteria for selection among such frames. The basic difficulty, the recognition of which inspires this study, is this: If facts are such only in a frame of reference, then we are confronted with a peculiar situation. In the first place no fact could occur which could cause a given frame of reference to be rejected. This is evident when we consider what a "fact" is. A fact, it is said, is the result of the interpretation of sense data by means of a conceptual scheme. The meaning of "conceptual scheme" is close enough to that of "frame of reference" to allow me to use the two interchangeably. A frame of reference involves a standpoint or a point of view from which phenomena are observed and in terms of which they are described. If the fact is the interpreted phenomenon, then clearly *only* those facts can occur which result when phenomena are described in the language of the given frame of reference. It becomes meaningless to say, as do Snygg and Combs, that science is seeking to sum up ". . . new facts inexplicable in the old frames of reference" (6). If no criteria exist for selecting among alternative frames of reference, then any frame of reference is of equal worth to any other frame and we have no way of deciding which to accept. Such a state of affairs would mean, e.g., that a Hullian frame of reference and a Freudian are matters of arbitrary choice. The suggestion at once arises, "Why not a complete relativism of such frames?" If the principle of relativity is to be applied, then we must be prepared also to accept the fundamental condition that makes relativity theory more than a mere babel of voices. The relativity principle in psychology may be stated in some such fashion as this: *The laws of psychology must be invariant under transformation from one system to another.* If such a condition is not accepted, then clearly chaos results. But here we are struck by another idea. If we accept the principle of invariance for psychological systems, does it not follow that these invariants must be expressed in some language, i.e., some frame of reference which, in a sense, must transcend all these other relative ones? We are then confronted with two problems different in nature: What are the invariants and what do they represent? The answer to this may

[1] Read to Southern Society for Philosophy and Psychology, Roanoke, Virginia, March 23, 1951.

lead us to a more precise definition of the subject matter of psychology than has hitherto been given. But for this paper, the more interesting question is this: Is this fact, the principle of invariance, itself meaningful in a frame of reference only? If it is, we are confronted by an infinite regress of frames of reference and we are involved either in a difficulty analogous to the logical problem of the theory of types in logic or to a kind of Gödel theorem. In both cases, we would be led to sterility in the controversy over frames of reference or to some form of unanswerable questions about frames of reference. The practical importance of these considerations may be indicated in the following manner. A good deal of work has been done to show the effect of frames of reference. So Asch (1) and others showed that judgments of traits like honesty, intelligence, etc. were a function of the individual's frame of reference. Even more vividly is this indicated in the study by Stagner (7) on Fascist attitudes in which a relation was established between economic status and pro-Fascist attitudes. Also Sherif (5) has established that the frame of reference tends to be established by the group. The implications of these experiments for the construction of theories and for the selection of frames of reference should be obvious. If the principles disclosed in these studies can be generalized to be applicable to scientific judgments, scientific opinions and scientific facts, then the implied relativism would make of the scientific enterprise merely a kind of "class-science" in a vicious form. No meaning could be given to the expression "objectivity of science" and no controversy that existed between frames of reference could be decided. Yet the question presents itself: In what frame of reference is the assertion of the relativism of frames of reference meaningful? In other words, it appears to me that the very assertion of the relativism of frames of reference involves the tacit assumption of some over-all frame of reference. From the point of view of theoretical psychology, it may be asked, when Sherif, e.g., points out the effect of frames of reference on perception, in what frame of reference is this study embedded, i.e., in what frame of reference are his perceptions of these effects meaningful? In the light of what I have said, the possibility arises that all facts about the effects of frames of reference on perception are meaningless in other frames of reference. If this possibility is to be avoided, then facts about frames of reference ought to be translatable into facts in other *acceptable* frames of reference.

In order to clarify my point, let me take two statements as examples:

1. What is perceived is determined by cultural conditions.
2. Neurotic behavior is the response of the individual to threats to his phenomenal self.

Although not direct quotations, it is apparent that the first might have been taken from Sherif's monograph, and the second from Snygg and Combs.

Each of these statements is constructed within a given frame of reference, using, therefore, the basic concepts and definitions of the respective frames. Since the second sentence contains the concept phenomenal self, which we shall assume is not present in Sherif's frame of reference, the second sentence is either meaningless in Sherif's frame of reference or it can be made meaningful by defining "phenomenal self" in Sherif's terms, if that is possible. Suppose it were not possible to define the concept in Sherif's terms? Snygg and Combs clearly mean to assert the *existence* of something which they call by this term. Therefore, if it is not possible to define this term in Sherif's

frame of reference, we would seem to be left with the following alternatives: (*a*) Sherif's frame of reference is too narrow since there exists an X which cannot be subsumed into it; (*b*) Snygg and Combs' frame of reference is too wide since it includes a term which cannot be embedded in Sherif's frame of reference. The question depends on the existence of a referent for a term. If the referent can be established as existing, then Snygg and Combs' frame of reference would appear to be preferable in this respect to that of Sherif's, because it includes a concept or term to denote that referent while Sherif's does not.

But this outcome immediately raises another question. Consider the statement: There exists a phenomenal self.

If we accept the notion that facts are such only within a frame of reference, and that statements are meaningful and true in a similar fashion, then the assertion of the existence of the phenomenal self is meaningful *only within the frame of reference of Snygg and Combs,* and meaningless in that of Sherif's *unless it can be translated into a statement in Sherif's frame of reference.* Therefore, the statement "a phenomenal self exists" cannot be used to refute Sherif's system. Furthermore, in Snygg and Combs' system, the statement is true *by definition* since the concept "phenomenal self" is so defined as to refer to certain already selected phenomena. Hence the assertion of the existence of the phenomenal self can never be rejected as long as one stays within the given system. If this is the last word, then clearly frames of reference are simply projections of the personality of the particular scientist—but even such an assertion is meaningful only in a frame of reference and the whole scientific task dies. The only way out is to recognize that an existence statement must have a reference beyond the bounds of the system in which it occurs and its verification is, in a sense, independent of any particular frame of reference. Failure to be able to translate an existence statement from one frame of reference to another would appear to leave only the following alternatives: Either the object asserted to exist does not exist, or the system in which its existence cannot be asserted is inadequate. In specific terms: If in Snygg and Combs we assert the existence of the phenomenal self, and this cannot be translated into an existence statement in Sherif's system, then either the phenomenal self does not exist, or Sherif's system is inadequate. In passing, we note that the failure to show the phenomenal self to exist does not mean a necessary rejection of the concept since clearly it may be a useful fiction or an intervening variable or a construct.

This part of our discussion leads, it seems to me, to a second area of invariance. Not merely must the laws of psychology be invariant under transformation from one system to another, but (*a*) *existence statements must transform into existence statements,* and (*b*) *if an existence statement is true in one system or frame of reference, it must be true in any other.* This second principle has rather startling implications. It means, for example, that a statement, e.g., in Freudian psychology such as: "the id exists," must transform into an existence statement in, e.g., behavioristic terms. That is why the book on *Personality and Psychotherapy* by Dollard and Miller (2) becomes important. If the statement "the id exists" is true in Freudian psychology, its transform must likewise be true in a behavioristic frame of reference.

These considerations lead to some elementary considerations. If a statement in a given system (call it S_1) must be transformable into a statement (S_2) in another system in such fashion that if

S_1 is true (or false), S_2 must be likewise, then this is possible if S_1 and S_2 are synonymous. But synonymity means equivalence of meaning or identity of referent. This implies, it appears to me, that the meaning of a statement must be clearly separated from its expression and is, therefore, independent of the frame of reference in which it is expressed. In order to take care of those cases where statements are about intervening variables or constructs, we have separated "meaning" and "referent." No one would, probably, confuse the referent with its description or with a statement about it.

With these thoughts in mind, we can say that by "phenomenon" we shall refer to what is meant or referred to. This would, then, lead us to say that science is concerned not with facts but with *phenomena*. Now, if S_1 is a statement in a given frame of reference about a given or even a presumed phenomenon, then it must be transformable into S_2, a statement in another frame of reference, in such a way that if S_1 is true (or false), then S_2 must also be true (or false). In particular, if S_1 is a statement of the form (*a*) "there exists an X, such that S_1 (X)," then this must be transformable into a statement of the form (*b*) "there exists an X, such that S_2 (X)." If (*a*) is true, (*b*) must be, and if (*a*) is false, (*b*) must also be. This means that we can now distinguish between true statements and false ones and use the word "facts" to refer to the former. Both "facts" and "non-facts" can occur in any system on the basis of these distinctions, and the task of the investigator is to be able to write down as many facts as possible. Also, a fact in one system must be transformable into a fact in another or else the system into which it cannot be stated must be rejected as inadequate at best and wrong at worst. *The invariants for both systems are the phenomena.*

It is necessary, therefore, to differentiate two aspects of any frame of reference: (*a*) its formal structure, and (*b*) its material signification. I shall not consider its formal structure because by now the form of a scientific system has been discussed at length and is relatively familiar to most investigators. Also, as is well known, the interpretation of the formal structure consists in the establishment of a set of rules of correspondence for fixing the meaning of the set of symbols. A system S_1 and another S_2 may have the same structure and may use the same or different symbols. They may use the same symbols to denote different phenomena, or different symbols to denote the same phenomena. When different symbols are used to denote the same phenomena, the task of transformation is easy. It amounts to establishing a dictionary of one frame of reference in terms of the other and S_1 transforms easily into S_2. Where the same symbols are used in S_1 and S_2 to denote different phenomena, then the relativism may be overcome by the simple expediment of constructing a system S_3, using a totally new set of symbols and establishing a dictionary for it in terms of S_1 and also another in terms of S_2. The statements in S_1 will translate into statements in S_3 which are then translatable into statements in S_2. S_1 and S_2 can be compared through S_3. Notice that this depends on being able to "know" S_1 and S_2, i.e., to be able to understand their meanings. Where different sets of symbols are used for different sets of phenomena, we may have two complementary systems. In every case the relativism must be transcended by reference to the phenomena.

There is one final set of considerations that needs to be recognized. A system S_1 may use, as a basic set of symbols, those referring to a given set of phenomena which are derivative in S_2. All

that has been said remains unaltered. The statements must be transformable within the systems. However, statements *about* the systems will differ, but with this we are not concerned. Which set of phenomena are to be symbolized as basic, depends on two considerations which are often confused: (*a*) metalinguistic ones, e.g., simplicity of system structure, fruitfulness, etc., and (*b*) material ones, e.g., whether as a matter of fact neuroses are reducible to repressions.

In conclusion, we point out that the important thing is the phenomena which give rise to facts, or which are what make statements into facts. It is the number of facts which one can derive that makes a system fruitful and applicable and not the number of statements. And while frames of reference are partially relative, the acceptance of a frame of reference is a function of its adherence to the principle of the relativity of systems—a principle which introduces an element of invariance and nonrelativity.

REFERENCES

1. ASCH, S. E., BLOCK, HELEN, & HERTZMAN, M. Studies in the principles of judgments and attitudes. *J. Psychol.*, 1938, 5, 219–251.
2. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
3. KATTSOFF, L. O. Observation and interpretation in science. *Phil. Rev.*, 1947, 56, 682–689.
4. KATTSOFF, L. O. The role of hypothesis in scientific investigation. *Mind*, 1949, 58, 222–227.
5. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, 27, No. 187.
6. SNYGG, D., & COMBS, A. W. *Individual behavior.* New York: Harper, 1949.
7. STAGNER, R. Fascist attitudes. *J. soc. Psychol.*, 1936, 7, 309–313.

[MS. received December 3, 1951]

Psychological Review
Vol. 60, No. 1, 1953

LEARNING AND THE SCIENTIFIC ENTERPRISE

DAVID BAKAN
University of Missouri

I

Whoever has given thought to the problems inherent in current views of learning phenomena might well wonder whether it is appropriate, at the present time, possibly to confound the issues by drawing attention to the fact that the scientific enterprise itself is a learning enterprise. In a sense, this point is banal. In another sense, it is not; for it is a point that has been largely ignored, and a point which has been systematically banished (2) by some theorists.

Fairly recently Skinner (7) has raised the question, "Are theories of learning necessary?" and has answered it in the negative. He has indicated that we can do better and faster without the types of learning theories which are currently available. In effect, what Skinner has done is to raise the controversy between empiricism and rationalism. But in doing so, Skinner also advances, by implication, a view concerning the nature of human learning. In effect, what Skinner is saying is that the attitude of empiricism is more conducive to efficient learning than is the attitude of rationalism. He makes an assertion about the role of attitude in human learning. This he does by advocating that the scientist shall concern himself primarily with the collection of data rather than the construction of elaborate theoretical systems. To quote him directly:

> An adequate impetus is supplied by the inclination to obtain data showing orderly changes characteristic of the learning process. An acceptable scientific program is to collect data of this sort and to relate them to manipulable variables, selected for study through a common sense exploration of the field (7, p. 215).

In advocating one "scientific" approach rather than another, one must necessarily be presupposing an hypothesis concerning the nature of human learning. The scientific enterprise is a learning enterprise. Thus any point of view which asserts what is good scientific method, presupposes a theory of human learning. The thesis of this essay is that if we were carefully to examine our views concerning the nature of the scientific enterprise, we might be in an excellent position to advance our understanding of the processes of human learning. In the same way that we were able to advance our understanding about the nature of human personality by the study of extremes, so might we be able to advance our understanding of the learning process by the study of an extreme in human learning, namely, the learning of the scientist.

There is a point of view in the field of psychology which sets up a "methodological" distinction between the scientist and the subject (2). "Methodologically" it differentiates between the man in his functioning as a scientist and in his functioning as a human being. This distinction is both unfortunate and unreal. Its maintenance closes off from our view one of the richest sources of information about human learning that we have. The distinction makes it impossible, in principle, if not in practice, to study the learning of the scientist in his learning operations. In advocating this distinction Bergmann and Spence (2) write:

> In the schema outlined by the scientific empiricist the experiences of the observing scientist do indeed have a privileged, even unique, position. If pressed too far and without the necessary epistemological so-

phistication, this account of the scientist's position can very easily lead to a metaphysical thesis of the solipsistic type (2, p. 3).

The nature of this "epistemological sophistication" is given to us by these writers as follows:

> . . . the empiristic scientist should realize that his behavior, symbolic or otherwise, does not lie on the same methodological level as the responses of his subjects . . . (2, p. 4).

Whatever "methodological" distinctions we may attempt to draw, we cannot escape from the brute fact that the scientist's "behavior, symbolic or otherwise" is behavior; and is, more particularly, learning behavior. We are hardly convinced that the acceptance of what the scientist does as behavior will lead us to solipsism, with the exception of one aspect of the solipsist position. The aspect of solipsism to which we have reference is its emphasis on the creative character of learning, a feature indicated by Tolman (10), who is hardly a solipsist. Indeed, the methodological distinction *preserves* rather than justifies a "privileged . . . unique position" for the experiences of the scientist. Thereby it grants creativity to the scientist, but denies it to the subject.

Not only does this distinction between scientist and subject cut off the opportunity of the examination of the scientific enterprise as an example of learning, but also tends to lead us into error in our investigations. In order to demonstrate this latter point, let us examine the notion that the stimulus is an independent variable. We take the notion of the stimulus as an example to point out one type of error that the "methodological" distinction can foster; and also we offer the analysis of the stimulus notion as an example of the type of analysis of the scientific enterprise which is being advocated.

II

The stimulus as an independent variable is a notion that is quite taken for granted by the "scientific empiricist." Spence (8) characterizes stimulus variables as follows:

> S-variables: measurements of physical and social environmental factors and conditions (present and past) under which the responses of organisms occur. These are sometimes referred to as the independent, manipulable variables (8, p. 48 f.).

The questions of whether the word "variable" applies, or the word "independent" applies, are not generally raised. Certainly the recent work on the effect of personality and attitude on perception can well lead us to ask whether we really know what we mean by a stimulus. A recognition of the difficulty inherent in this matter of perception is manifested in the effort to reconcile the lack of independence of the stimulus with Hullian theory (3).

What do we mean by a variable? Let us consider two items, the status of which as variables would hardly be challenged, say, length and color. By virtue of what may these be considered to be variables? First, that they vary, i.e., that there are two or more lengths, and that there are two or more colors. Second, that the designated categories of the variables shall be mutually exclusive, i.e., if something be of one length, it not be of another length, and if something be of one color, it not be of another color. That which distinguishes one variable from another is that the criterion of mutual exclusiveness does not apply, i.e., if something be of one color it may be of any length, and if something be of one length, it may be of any color. Thus we arrive at a definition of a variable:

A variable is a set of two or more categories such that, if any object or event be a member of one of those cate-

gories, it may not be a member of any other of those categories.

Let us consider some of the implications of this definition for the notion of a stimulus variable. Let us avoid getting into the metaphysical question by simply saying that the real world is what it is, without any further elaboration. However, whatever the nature of the real world may be, and without in any way denying the existence of the real world, the *constitution*[1] of the variable *as a variable* is the work, or invention, or creation, of the scientist. Variables, whatever they may be *in re,* do not exist there *as variables.* For, variables are, by definition, sets of categories; and *categories are the result of someone's delineation, abstraction, and identification.*

By virtue of these considerations, if it were not so tedious, it would be more appropriate to talk of variables-as-constituted-by-the-scientist, rather than simply of variables. The refinement that is introduced by this distinction is less important in the physical sciences, since, in the latter, it would be rather meaningless to talk of variables-as-constituted-by-the-subject. Even in the physical sciences, following the notions of Bridgman (4), the distinction is appropriate. However, this distinction is extremely important in the psychology of learning, for here we have the constitution of variables both by the experimenter and by the subject.

To clarify this point, let us consider a more or less typical experimental situation in the psychology of learning. Consider a Skinner box, with a lever, the depression of which by the animal results in the delivery of a food pellet. The animal is placed in the box. Is there any constitution of the stimulus variable on the part of the animal? Not at the outset of the experiment. However, the lever is constituted at the outset of the experiment as a stimulus variable by the experimenter. (The categories of the variable in this case are presence or absence of the lever.) It is *delineated* from the remainder of the environment, the property of leverness has been *abstracted,* and it is *identified* as a lever, the depression of which leads to the delivery of a food pellet, *by the experimenter* at the outset of the experiment.

As "learning" takes place or proceeds, the animal comes to depress the lever with greater and greater regularity. The animal, it may be said, comes to learn *what's what.* We prefer "what's what" to the Tolmanian "what leads to what" since the former, to us, seems to include the delineation and the abstraction more than the latter does. The animal delineates the lever from the remainder of the environment, avails itself of its leverness, and identifies it as leading to food. Essentially, what takes place in this type of experimental situation is that the animal comes, after the learning process, to constitute the stimulus variable in much the same way as the experimenter did at the outset of the experiment.

The usual experimental learning situation is a teaching situation, more particularly, than a learning situation. The *success* of the learning is indicated by the degree to which the animal seems to have constituted the stimulus variable as the experimenter has constituted the stimulus variable. The latter is exactly what is meant by the *learning criterion* in most experiments in the psychology of learning.

The "methodological" distinction between scientist and subject prevents us from seeing that the *psychological processes of the experimenter are an integral*

[1] We borrow this word from Dewey's *The reflex arc concept in psychology* (5). Much of what is being said here is based on our interpretation of some of the notions in Dewey's paper.

part of the nature of the experimental situation. Furthermore it leads us into the error of generalizing from such experimental situations to learning in general, when actually it is the special case of learning which we might call teaching or communication. "Methodologically" it removes the experimenter from the "data" when actually his psychological processes are "data" too; and necessary data for the understanding of the nature of the experimental situation.

III

We have concerned ourselves with the *variable* aspect of the stimulus variable. Let us now turn our attention to the question of the *independence* of the stimulus variable. What is generally meant by the assertion of the independence of the stimulus is that it is independent of the response. The paradigm is that the response is dependent on the stimulus, but the stimulus is independent of the response. The formulation of this paradigm is

$$R = f(S).$$

But even here, the matter is begged. For, in the explanation of what a stimulus variable is, the expression "under which the responses of organisms occur" (8) is used. The stimulus does not exist as a stimulus except by virtue of the responses of the organism. The delineation, abstraction, and identification of the stimulus *is the response.*

The only kind of independence that the stimulus variable can have, is a spurious one. The delineation, abstraction, and identification of the stimulus variable *by the experimenter* is independent of the delineation, abstraction, and identification of the stimulus variable *by the subject.* But in the subject itself, there is no such independence.

The operationist argument to the effect that our knowledge is contingent upon the operations that we engage in in obtaining this knowledge can well be leveled against the conception that the stimulus variable is independent of the response. For we cannot even identify the stimulus properly without taking account of the responses of the organism. If this is true, we can hardly talk of the independence of the stimulus variable.

Again, we might point out that the fallacy of considering the stimulus variable as independent may be viewed as stemming from the "methodological" distinction between scientist and subject. It can be said that, by virtue of not taking account of the psychological processes of the investigator, the investigator is led into a position where he "projects" his prior constitution of the stimulus into the organism that is the subject in the experiment.

IV

The point which has been made above, that it is inappropriate to talk of the independent stimulus variable, has, interestingly enough, been made in current discussions of the scientific method by some of the same persons who talk of the independent stimulus variable. The point is one of the essential ones of operationism. It has, however, not been brought to bear as an insight concerning learning. Thus, for example, Bridgman (4) has given priority, not to the stimulus, but to the response. Basic to scientific learning is not the stimulus, but the *operations of the scientist.* Similarly, Stevens (9), accepting the priority of the operation, goes on further to specify the nature of the operation as being *discrimination*, a response. Bergmann and Spence (1), aside from their particular views on the psychology of learning proper, have emphasized the response feature of the scientist's learning, in their emphasis on theory construction and theory creation. In spite of their avowed point of view concerning the

distinction between scientist and subject, they still permit themselves to say:

> Historically and *psychologically*, then, the *creation* of helpful concepts is a very essential part of a scientific achievement (1, p. 3) (our italics).

and:

> Actually much of what is usually called theorizing in empirical science consists . . . in the *creation* of these organizing empirical constructs . . . (1, p. 6) (our italics).

In these writings we find not only a stress upon the response priority in the learning situation, but also the creative aspects which are involved in learning. We have here a stress upon a feature of human learning which is rarely mentioned in the works on the psychology of learning proper. It is of more than parenthetical interest that Hull's *Principles of Behavior* (6), an effort in human learning, is certainly worthy of the adjective "creative."

V

There is certainly much more to be learned about the nature of human learning by the consideration of the scientific enterprise as an instance of human learning. Here we have made no more than the attempt to suggest where this notion might lead, and to outline a program in broad brush strokes. In the scientific enterprise we have an example of learning *de novo* as contrasted with learning by being taught. By the study of it we might learn a great deal concerning the nature of learning.

REFERENCES

1. BERGMANN, G., & SPENCE, K. W. Operationism and theory in psychology. *Psychol. Rev.*, 1941, **48**, 1–14.
2. BERGMANN, G., & SPENCE, K. W. The logic of psychophysical measurement. *Psychol. Rev.*, 1944, **51**, 1–24.
3. BERLYNE, D. E. Attention, perception and behavior theory. *Psychol. Rev.*, 1951, **58**, 137–146.
4. BRIDGMAN, P. W. *The logic of modern physics.* New York: Macmillan, 1938.
5. DEWEY, J. The reflex arc concept in psychology. *Psychol. Rev.*, 1896, **3**, 357–370.
6. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
7. SKINNER, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, **57**, 193–216.
8. SPENCE, K. W. The nature of theory construction in contemporary psychology. *Psychol. Rev.*, 1944, **51**, 47–68.
9. STEVENS, S. S. Psychology and the science of science. *Psychol. Bull.*, 1939, **36**, 221–263.
10. TOLMAN, E. C. Theories of learning. In F. A. Moss (Ed.), *Comparative psychology.* New York: Prentice-Hall, 1934.

[MS. received December 3, 1951]

Psychological Review
Vol. 60, No. 1, 1953

VALUES, WORD FREQUENCIES, AND PERCEPTION

JOE ADAMS AND DONALD R. BROWN

Bryn Mawr College

In a recent article Solomon and Howes (11) have made two points about the research by Postman, Bruner, and McGinnies (7) supposedly demonstrating a relationship between value and perception:

1. Since duration thereshold (i.e., duration of tachistoscopic exposure necessary for recognition to occur) is related to word frequency,[1] the relation of duration threshold to value rank on the Allport-Vernon test (1) can be accounted for by assuming that "for words representing a field of interest, valuation of that interest is associated with a departure from the mean frequency of use in the general population" (11, appendix).

2. The Allport-Vernon test itself can be considered a measure of the frequency with which the subject uses certain words—and nothing more than that; that is, there is no need to postulate or infer any vague entities like "values" in order to account for value ranks obtained when the test is administered to a subject. The authors make this point quite explicit by postulating two hypothetical populations A and B which would obtain different value ranks on the test even though "the only difference between the two populations is the frequency with which they use the two sets of words" (11, p. 265).

In replying to Solomon and Howes, Postman and Schneider (8) merely gather additional data demonstrating the relation between value rank (as measured by the Allport-Vernon test) and perceptual-cognitive functions (visual recognition and recall). They obtain significant differences between duration thresholds for both relatively high- and relatively low-frequency words (in the general population) as a function of value rank. Though they disagree with Solomon and Howes with regard to the question of hypothetical cognitive and motivational processes versus stimulus response correlations and response probabilities without such hypothetical constructs, the fact remains that their results can be predicted on the basis of point 1 above plus a slight additional assumption about the relation of word frequency to recall.

We do not wish to take issue with point 1 made by Solomon and Howes; it seems plausible enough and should be taken into account by anyone doing research on the relation of motivational processes or cognitive structures to duration thresholds, when words are used as stimuli. This point should probably be modified, however, if, as it seems to us, the Allport-Vernon test confounds to some extent two psychological dimensions which can be separated, namely *interest* and *value*. An individual can be interested in a given area even though he has a strong disagreement with individuals or institutions operating in that area. For example, a militant atheist may be very interested in religion though harboring little value for religious beliefs or experience. Because of the way in which the Allport-Vernon test is constructed and scored, it seems to us that interest and value are confounded, though no doubt these two variables are

[1] This relation was found by Howes and Solomon (4) to be logarithmic, i.e., $t = a \log f + b$, in which t is duration threshold, f is word frequency, and a and b are constants. The scatterplots which the authors present, however, seem to indicate a nonlinear relation between t and $\log f$.

correlated to some extent. It is reasonable to believe that whatever correlation exists between word frequency and value rank on the test can be accounted for more by the interest variable than the value variable; at any rate this hypothesis could be tested. Of course our formulation will not be acceptable to Solomon and Howes; if they wish to make this distinction they will have to redefine "interest" and "value" in accordance with their own theoretical orientation (see below).

With regard to point 2 we wish to point out the following: (*a*) The analysis by Solomon and Howes of what happens when a subject takes the Allport-Vernon test is somewhat unsound; (*b*) the Allport-Vernon test is inaccurately described by Solomon and Howes, even with respect to the mechanical features of taking and scoring the test.

Let us first describe roughly the two interpretations given by Solomon and Howes of what happens when a subject takes the Allport-Vernon test. As the reader may have the impression that we are misrepresenting these interpretations, we refer him to the careful exposition by Solomon and Howes on pages 265 and 266 under the heading "The Allport-Vernon test as a corollary of word frequency." First of all, Solomon and Howes assert that the test can be considered as a "set of visual choice discriminations between alternative groups of words, the choice being presented as a series of paired comparisons. Each comparison is between words representing two different interest areas. For example, consider item 4 of the test: 'If you were a university professor and had the necessary ability, would you prefer to teach: (*a*) poetry; (*b*) chemistry and physics?'" Solomon and Howes assert that if a subject were to indicate his choice by responding vocally, then, if he had a lower duration threshold for "poetry" than for "chemistry and physics," he would have a tendency to make the response "poetry," because "on this test the operations performed are identical to those defining duration threshold" (11, p. 265). They then assume "the response of checking a number that corresponds to a printed word does not differ fundamentally from the response of emitting the vocal sounds" (cf. Weiss (14, pp. 301–304)). Since duration threshold has been shown to be related to word frequency, the Allport-Vernon test is thus reduced to a corollary of word frequency.

It seems to us that the operations performed in giving and taking the Allport-Vernon test are not identical with those defining duration threshold. If the subject were asked to read the first part of the item, then glance at the remainder and check the first word that he sees, then perhaps the operations would be highly similar, though such methodological controls as fixation point, etc., would be absent. Furthermore, when Solomon and Howes quote their own experiment, "we know from (4) that, following any given amount of exposure to the printed alternatives, the word with the higher frequency of use will tend to occur more often," they overlook the fact that in their experiment (4) this relationship holds over only a very short range of time exposures, and that unless they can show that the time of exposure of the alternatives on the Allport-Vernon test is within this range, they are extrapolating their results in a most extraordinary way.

The second interpretation of what happens when a subject takes the Allport-Vernon test runs somewhat as follows: the answer indicated by the subject by "circling" an answer is determined by the strength of association between the word (visual stimulus) and the response (circling); "the printed words *poetry* and *chemistry or physics* are the discriminative stimuli and the

circling of an answer is the overt response in this discrimination. In conventional association interpretation this response is assumed to be mediated by the covert response of one of the alternative word-pairs, and the appropriate circling response will depend upon which mediating response is more strongly associated with its visual stimulation" (11, p. 266). If this interpretation is correct, then perhaps the first part of the item can be dispensed with. We are willing to predict that dispensing with the first part of the item will give somewhat different results.

Our second point may make the preceding discussion seem rather academic. The subject does not indicate his answer on the Allport-Vernon test by checking or circling a number, letter, or word. We quote from the directions of the test (1) and present the first item:

> A number of controversial statements or questions with two alternative answers are given below. Indicate your personal preferences by writing the appropriate figures in the right-hand columns, as indicated:

	(*a*)	(*b*)
If you agree with alternative (*a*) and disagree with (*b*), write 3 in the first column and 0 in the second column, thus	3	0
If you agree with (*b*); disagree with (*a*), write	0	3
If you have a slight preference for (*a*) over (*b*), write	2	1
If you have a slight preference for (*b*) over (*a*), write	1	2
	(*a*)	(*b*)
1. The main object of scientific research should be the discovery of pure truth rather than its practical applications. (*a*) Yes; (*b*) No.	—	—

Items 2 and 3 are also of the Yes-No variety, as are 12 items in all of the 30 items in the first part of the test.

Part II of the test has somewhat different directions from those of Part I:

> Each of the following situations or questions is followed by four possible attitudes or answers. Arrange these answers in the order of your personal preference from first to fourth by writing, in the left-hand margin:
>
> ..1.. beside the answer that appeals to you most,
> ..2.. beside the answer which is next most important to you,
> ..3.. beside the next, and
> ..4.. beside the answer that least represents your interest or preference.
>
> 1. Do you think that a good government should aim chiefly at
>
> ..... *a.* more aid for the poor, sick, and old,
> ..... *b.* the development of manufacturing and trade,
> ..... *c.* introducing more ethical principles into its policies and diplomacy,
> ..... *d.* establishing a position of prestige and respect among nations.

It would appear, from a quick inspection of the Allport-Vernon test directions and items, that the choice involves something more than a simple paired comparison, not to mention the fact that a significant part of one's value score seems in Solomon's and Howes's terms to be dependent upon one's fortuitous experiences with the words "Yes" and "No."

However, we have constructed a modified Allport-Vernon test in which a circling response is made to test Solomon and Howe's hypothesis about the test as they conceive it.

We should now like to make a few general remarks about the theoretical orientation of Solomon and Howes. These authors attempt to rid psychology of vague mentalistic terms like "perception" (as ordinarily used) by defining "seeing," "perception," and other vague mentalistic terms operationally by care-

ful specification of stimulus and response properties and their publicly observable relationships. This orientation is shared by some other psychologists (2, 3, 5, 6, 10, 12, 13, 14). Solomon and Howes speak of "optical situations" and "visual choice discriminations" (e.g. as in the Allport-Vernon testing situation), "linguistic responses," "circling responses," etc., implying that a great gain in scientific advance in psychology would result were other psychologists to do likewise. The operationist-nonoperationist, private-public controversy is one which we certainly do not wish to attempt to discuss in detail in this paper. The argument is in part a semantic one, because any psychologist, whether he *defines* "perception," etc. operationally or not, has to develop operational *indices* of what he wants to study whenever he uses other individuals as subjects. We might call the first an "Operationist" and the second an "operationist." Many psychologists who are operationists are not Operationists. The controversy is then in part a controversy about whether observables are to be used as indices of something else or whether the observables are to be the defined subject matter of investigation. This controversy is not entirely semantic, however, because it is in part a controversy about (*a*) whether phenomenology is an acceptable method, (*b*) what terms psychologist should use in their "thinking" so as to advance psychology as a science.

The authors believe that psychologists can sometimes obtain valid knowledge (not mere *hints* about how to proceed with other subjects) by observing contrast effects, after-images, apparent movement, figural after-effects, etc. We also believe that it sometimes pays a psychologist when talking about certain phenomena, e.g., after-images, to think in terms of vague, mentalistic events, like after-images, rather than stimulus-response correlations (cf. Pratt, 9). Further, we challenge Operationists to prove by an examination of the history of science or any other method that "Operational" thinking is in all cases—or even in most cases—the most advantageous way of thinking. We do not believe, moreover, that certain Operationists have freed themselves from mentalistic ways of thinking. It seems to us that if Solomon and Howes were consistent in thinking of the Allport-Vernon test in terms of "optical operations" and "circling responses," they would, being good psychologists, have been careful to control such conditions as fixation point, time of exposure, position of the pencil and arm on the table, etc., and perhaps they would have carefully distinguished between checking and circling as different motor responses.

It would be possible to make a direct test of Solomon and Howes's frequency hypothesis by altering the Allport-Vernon test in several ways. Two methods that immediately suggest themselves would be to put all the statements in a negative way by the use of "not," "rarely," "hardly ever," or "never." Presumably, then, the subject would have to resist his "circling responses" to the high-frequency valued words and make the response contrary to frequency. We have revised the test in such a way that the alternatives are all in very infrequent synonyms except those scoring in one value area which are phrased in very frequent words. Six such forms of the test have been constructed, each having one of the value areas favored frequencywise. Presumably the subject would be duped into the high-frequency words and thus the value area so designated by the experimenter.

REFERENCES

1. ALLPORT, G. W., & VERNON, P. E. *A study of values.* Boston: Houghton-Mifflin, 1931.
2. BORING, E. G. The use of operational definitions in science. *Psychol. Rev.,* 1945, 52, 243–245.

3. GRAHAM, C. H. Behavior, perception, and the psychophysical methods. *Psychol. Rev.*, 1950, 57, 108–120.
4. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, 41, 401–410.
5. MEYER, M. *Psychology of the other-one.* Columbia, Mo.: Missouri Book Co., 1921.
6. MILLER, G., & SELFRIDGE, J. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, 63, 176–185.
7. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, 43, 142–155.
8. POSTMAN, L., & SCHNEIDER, B. H. Personal values, visual recognition, and recall. *Psychol. Rev.*, 1951, 58, 271–284.
9. PRATT, C. C. *The logic of modern psychology*. New York: Macmillan, 1939.
10. SKINNER, B. F. The operational analysis of psychological terms. *Psychol. Rev.*, 1945, 52, 270–277.
11. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, 58, 256–270.
12. STEVENS, S. S. Psychology and the science of science. *Psychol. Bull.*, 1939, 36, 221–263.
13. WATSON, J. B. *Behaviorism.* New York: Norton, 1930.
14. WEISS, A. P. *A theoretical basis of human behavior.* Columbus, Ohio: R. G. Adams, 1950.

[MS. received December 17, 1951]

Psychological Review
Vol. 60, No. 1, 1953

PRELIMINARY SUGGESTIONS AS TO A FORMALIZATION OF EXPECTANCY THEORY[1]

KENNETH MACCORQUODALE AND PAUL E. MEEHL

University of Minnesota

In the present paper we shall try to indicate one direction in which a rigorization of expectancy theory might move, concentrating wholly upon what we believe are the major constructs of such a theory. In a previous paper (9), we discussed briefly the question, "How may an expectancy theory of learning be identified as such?" We took the position that several features in the thinking of expectancy theorists (e.g., Tolman) are not logically entailed by the admission of an expectancy construct, and we suggested that one (and perhaps *the*) crucial differentiator between an expectancy and a nonexpectancy theory is the form of the acquisition postulates. One particular type of acquisition postulate, providing for the strengthening of the basic learning element in a special way, we called the "Inference Postulate," listed as Number 4 below.

While the presence of such a postulate generates many properties in the theory which are absent without it, strictly speaking it is always the entire *system* of postulates which does the generating. The inference postulate, for example, cannot lead to the desired deductions concerning latent learning unless the role of the expectancy construct in activating behavior is also indicated (our postulate 12). This latter indication in turn requires some subset of postulates defining motive-incentive constructs (as our postulates 6, 7, 9, 10, 11), and so it goes. Ideally, the implicit definition of these constructs would be given by the entire, "complete" postulate system; their empirical meaning would be, so to say, exhibited, *shown forth* by the interdigitation of the propositions with one another and with the behavioral theorems they jointly entail.

In what follows we shall present an *incomplete*, *tentative*, and certainly *nonsufficient* set of propositions which, however, begin to define at least one important kind of expectancy theory. This formulation has a major affinity with that of Tolman because it contains an inference postulate, and (what in turn makes *this* possible) because its fundamental cognitive unit, the "expectancy" (S_1RS_2) is designated by three notational elements rather than the two which specify Hull's $_SH_R$. In this set, then, the basic construct of learning specifies not only the S_1 in the presence of which the organism emits R, but in addition it specifies what is expected when R follows S_1.

On the other hand, this formulation differs from Tolman's in several respects; of these differences we currently consider only one to be clearly fundamental. The fundamental difference is our inclusion of an R *reference* in the notation specifying the basic cognitive unit, the "expectancy." In this sense the construct lies somewhere between the constructs of Tolman and those of Hull, since it includes the S_2 reference (unlike Hull)

[1] This paper grew out of discussions at the Social Science Research Council-subsidized Dartmouth Conference on Learning Theory (1950). Although they are in no sense to be considered as endorsing our formulations, the other members of the Conference—Drs. W. K. Estes, C. G. Mueller, S. Koch, W. N. Schoenfeld, and W. S. Verplanck—contributed both stimulation and criticism.

but also the *R* reference (unlike Tolman). It is true that in some of his discussions (e.g., **14**, pp. 10–12, 82) Tolman seems to include the response reference in his idea of an expectancy; but his more recent emphasis on perception and the "map" metaphor have deflected attention from the *R* term. This sort of emphasis leads him to have some difficulty in *getting to behavior*, which is expressed in Guthrie's well-known gibe about Tolman's leaving the rat "buried in thought." We try to solve the problem by including an *R* reference as part of the expectancy construct from the beginning. Let us emphasize, however, that this *must not be understood to mean that an expectancy IS a "response"* if the word response means an effector-event-class; nor does it mean that the arousal of an expectancy is an effector-event, albeit a damped or attenuated one.

The facets of Tolman's approach which we do not see as crucial or differentiating were merely listed in our previous paper (**9**, p. 230). Professor Tolman informs us (personal communication) that he agrees with us about four of them, but he feels that two of them, the "Gestalt-configural stress" and the "specification of re-action-class by reference to position, direction, or locomotion (rather than by effector properties)," are integral features of his view. Space does not permit us the detailed consideration of these two points which would be necessary to do them justice, but this whole question will be treated in detail in a subsequent publication.

We cannot stress too strongly that the following is *not* offered as a full-fledged "postulate set." The postulates represent our first attempts to nail down the expectancy view. They are obviously incomplete even if the theorem system required were a "miniature" one, e. g., confined to rat behavior in the Skinner box. The great mass of data regarding temporal effects, for instance, is left quite untreated. The notion of "similarity" between two stimulus configurations is left unclarified; there is no attempt to treat the entire area of facts and concepts usually called "inhibition," or the allied phenomena of non-extinctive work decrement. The postulate of need strength is not clearly supported by current evidence on alimentary drives and there is no good evidence that for other "higher-order" need variables it should be expected to hold at all. There are no postulates regarding the very important question of interaction among expectancies to yield some compromise *R* strength. We have not felt it propitious even to make guesses about the exact *form* of functions, at this point. But anyone who wishes to substitute, e.g., "simple positive growth function" wherever we have written "increasing decelerated function" can easily do so.

Since one of the constantly reiterated complaints against expectancy theorists for almost twenty years has been the extreme sketchiness of their formulations, we perhaps do not need to apologize for the present effort if it succeeds in reducing this sketchiness even slightly.

Furthermore, the present article aims merely to *present* the incomplete set and to *illustrate* its workings by deriving some semiquantitative consequences. No attempt is made *to defend* our particular decisions as to formulation, or to develop the thinking behind each. This more extended treatment will appear in a subsequent publication.

Pending detailed treatment, we shall merely say that in what follows the term *stimulus* may include a physical situation of any describable complexity or patterning. That is,

if "triangularity" is the property needed to yield a response-inferred stimulus equivalence, the relational features between three points are, of course, included in the specification of S. It is not suggested that there are no unsolved problems here; but since, like Skinner, we are unconvinced of the definitive role of "configural properties" as marking out kinds of learning theories, we have ignored these issues for present purposes. In the same way, we mean by *response* a class of effector activities which produce environmental effects within a specified range of values, e.g., "turning left," "pressing lever," and the like. (Cf. **5**, pp. 95–96; **12**, pp. 33–43.) Again, this summary treatment springs from an acute awareness of the terrible difficulties in the response concept rather than a naive belief that they have been solved.

The postulates are as follows:

1. *Mnemonization*: The occurrence of the sequence $S_1 \rightarrow R_1 \rightarrow S_2$ (the adjacent members being in close temporal contiguity) results in an increment in the strength of an expectancy $(S_1R_1S_2)$. The strength increases as a decelerated function of the number of occurrences of the sequence. The growth rate is an increasing function of the absolute value of the valence of S_2. If the termination by S_2 of the sequence $(S_1 \rightarrow R_1)$ is random with respect to nondefining properties of S_1, the asymptote of strength is $\leq$ the relative frequency P of S_2 following $S_1 \rightarrow R_1$ (i.e., a pure number). How far this asymptote is below P is a decelerated function of the delay between the inception of R_1 and the occurrence of S_2.

2. *Extinction*: The occurrence of a sequence $(S_1 \rightarrow R_1)$ if not terminated by S_2 produces a decrement in the expectancy if the objective S_2-probability has been 1.00, and the magnitude of this decrement is an increasing function of the valence of S_2 and the current strength of $(S_1R_1S_2)$. Such a failure of S_2 when P has been $\doteq 1$ is a *disconfirmation* provided $(S_1R_1S_2)$ was nonzero. If the objective probability P shifts to a lower P', and remains stable there, the expectancy strength will approach some value $\leq P'$ asymptotically.

3. *Primary Generalization*: When an expectancy $(S_1R_1S_2)$ is raised to some strength, expectancies sharing the R and S_2 terms and resembling it on the elicitor side will receive some strength, this generalized strength being a function of the similarity of their elicitors to S_1. The same is true of extinction of $(S_1R_1S_2)$.

4. *Inference*: The occurrence of a temporal contiguity S_2S^* when $(S_1R_1S_2)$ has nonzero strength, produces an increment in the strength of a new expectancy $(S_1R_1S^*)$. The induced strength increases as a decelerated function of the number of such contiguities. The asymptote is the strength of $(S_1R_1S_2)$ and the growth rate is an increasing decelerated function of the absolute valence of S^*. The presentation of S_2 without S^* weakens such an induced expectancy $(S_1R_1S^*)$. The decrement is greater if the failure of S^* occurs at the termination of the sequence $S_1 \rightarrow R_1 \rightarrow S_2$ than if it occurs by presentation of S_2 not following an occurrence of the sequence.

5. *Generalized Inference*: The occurrence of a temporal contiguity S_2S^* produces an increment in the strength of an expectancy $(S_1R_1S^*)$ provided that an expectancy $(S_1R_1S'_2)$ was at some strength and the expectandum S'_2 is similar to S_2. The induced strength increases as a decelerated function of the number of such contiguities. The asymptote is a func-

tion of the strength of $(S_1R_1S'_2)$ and the difference between S_2 and S'_2. The growth rate to this asymptote is an increasing decelerated function of the absolute valence of S^*.

6. *Secondary Cathexis:* The contiguity of S_2 and S^* when S^* has valence $|V|$ produces an increment in the absolute cathexis of S_2. The derived cathexis is an increasing decelerated function of the number of contiguities and the asymptote is an increasing decelerated function of $|V|$ during the contiguities, and has the same sign as the V of S^*. The presentation of S_2 without S^*, or with S^* having had its absolute valence decreased, will produce a decrement in the induced cathexis of S_2.

7. *Induced Elicitor-Cathexis*: The acquisition of valence by an expectandum S_2 belonging to an existing expectancy $(S_1R_1S_2)$ induces a cathexis in the elicitor S_1, the strength of the induced cathexis being a decelerated increasing function of the strength of the expectancy and the absolute valence of S_2.

8. *Confirmed Elicitor-Cathexis*: The confirmation of an expectancy $(S_1R_1S_2)$, i.e., the occurrence of the sequence $(S_1 \rightarrow R_1 \rightarrow S_2)$ when $(S_1R_1S_2)$ is of nonzero strength, when S_2 has a certain valence, produces an increment in the cathexis of the elicitor S_1.

This increment in the elicitor-cathexis by *confirmation* is greater than the increment which would be *induced* by producing a valence in S_2 when the expectancy is at the same strength as that reached by the present confirmation.

9. *Valence*: The valence of a stimulus S^* is a multiplicative function of the *need* D and the *cathexis* C^*. (Applies only to cases of positive cathexis.)

10. *Need Strength*: The need (D) for a cathected situation is an increasing function of the time-interval since satiation for it.

Note: Upon present evidence, even basic questions of monotony and acceleration are unsettled for the alimentary drives of the rat, let alone other drives and other species. There is no very cogent evidence that all or even most "needs" rise as a function of time since satiation, although this seems frequently assumed. Even the notion of satiation itself, in connection with "simple" alimentary drives, presents great difficulties. This proposition can, therefore, hardly be taken as having any generality even tentatively.

11. *Cathexis*: The cathexis of a stimulus situation S^* is an increasing decelerated function of the number of contiguities between it and the occurrences of the consummatory response. The asymptote is an increasing function of the need strength present during these contiguities. (There may be some innately determined cathexes, however.)

12. *Activation*: The reaction potential $_sE_R$ of a response R_1 in the presence of S_1 is a multiplicative function of the strength of the expectancy $(S_1R_1S_2)$ and the valence (retaining sign!) of the expectandum. There are momentary oscillations of reaction potential about this value $_sE_R$, the frequency distribution being at least unimodal in form. The oscillations of two different $_sE_R$'s are treated as independent, and the response which is momentarily "ahead" is assumed to be emitted.

As a sample of "derivation" from these, let us consider first a study by Kendler (6) which yielded somewhat puzzling results from the Hullian point of view. Rats were run in a single T maze when both hungry and thirsty. On one side the goal box always contained food, on the other water, and in both boxes the consum-

matory response occurred repeatedly during the first phase. On the test run, some animals were made thirsty, the others hungry. "Appropriate" choices were made to a pronounced extent. Kendler points out that since both left and right choices had been consistently reinforced during the training runs, and since both hunger and thirst were present as cue variables for choices in both directions, it is not clear on Hullian principles why the differential choice is shown on the one-drive test runs. He discusses two possibilities, one in terms of the fractional goal response (suggested by Spence) and the other involving a rather radical modification of the usual Hullian reinforcement principles, such that the only drive cues which get connected to R are those which are reduced by the reinforcing operation that strengthens the ${}_SH_R$ in question (Guthrie?). The finding would be treated within the present frame as follows:

S_C: Choice-point stimulation
R_R: Right turn (leads to food)
R_L: Left turn (leads to water)
S_{RF}: Stimulation in right goal box, including that of eating
S_{LW}: Stimulation in left goal box, including that of drinking

Then during the first phase we have the sequences:

$$S_C \rightarrow R_R \rightarrow S_{RF}^*$$
$$S_C \rightarrow R_L \rightarrow S_{LW}^*.$$

Repetitions of these in balanced amounts lead to increased strengths of two expectancies

$$(S_CR_LS_{LW}) \doteq (S_CR_RS_{RF})$$

according to (1). Unless the two drives are unequal or the number of exposures unbalanced, these two expectancies will rise at the same rate; toward the end of the first phase they should, on the average, be equal in cumulated strength. Assuming hunger and thirst drives to be equal, at this stage, by (10), (11), and (12), we have

$$\begin{aligned} {}_SE_{R(L)} &= (V_W)(S_CR_LS_{LW}^*) \\ &\doteq (V_F)(S_CR_RS_{FR}^*) = {}_SE_{R(R)} \end{aligned}$$

so no "preference" is manifested. On the test run, we satiate for food and keep the animals thirsty. That is,

$$(V_W) \gg (V_F)$$

so that

$$\begin{aligned} {}_SE_{R(L)} &= (V_W)(S_CR_LS_{LW}^*) \\ &\gg (V_F)(S_CR_RS_{RF}^*) = {}_SE_{R(R)}. \end{aligned}$$

By (12) the probability of a left turn is then much higher than for a right. It is instructive to ask where we find the *locus* of the difference that generates the derivation for us more readily than for Kendler. It evidently lies in the fact that *our activation postulate makes reference to the expectandum*, and this reference mediates a "control" over the strength of R which cannot be readily talked about in a notation designating the basic cognitive element by reference to S_1, the elicitor, and R only. Thus, the acquisition phase in Hullian terms is strengthening ${}_{S(C)}H_{R(R)}$ and ${}_{S(C)}H_{R(L)}$ equally, since turns both ways are reinforced. When it comes to predicting the test run, manipulating drive can only affect a multiplier of these ${}_SH_R$ elements, and the "forward-pointing" reference of the expectandum term is lacking or must be smuggled in by invoking r_g. In short, the Hullian frame makes the incentive important only "historically," i.e., in determining how much ${}_SH_R$ is cumulated. The expectancy frame retains the reference to the expectandum in the basic cognitive element (S_1RS_2), and thus a reference to this third thing can be packed into

the activation postulate in a way that allows subsequent drive manipulations a ready control over the response strength.

Consider next the experiment reported by Tolman and Gleitman (15). After a first phase in which they ran either to left or right to readily discriminable goal boxes, in each of which they were allowed to eat, the rats were then *placed* (not following a run) in one of the goal boxes where they were allowed to eat. Each rat was also placed in the other goal box, where a shock was administered. On the test run the rats chose appropriately to a very striking extent. Suppose shock is administered in the left-hand goal box during Phase II, and food given in the right. Let the entire stimulus complex involving eating food be designated $S_F{}^*$ and that involving shock by $S_{Sh}{}^-$. Then the reported effects could arise as follows:

I. Status at end of Phase I:

a. $(S_C R_L S_L) \doteq (S_C R_R S_R)$ By (1)

b. $(S_C R_L S_F{}^*) \doteq (S_C R_R S_F{}^*)$ By (1)

c. $V_{S(L)} \doteq V_{S(R)}$ By (6)

II. Status at end of Phase II:

d. $(S_C R_L S_L) \doteq (S_C R_R S_R)$

e. $(S_C R_L S_F{}^*) < (S_C R_R S_F{}^*)$ By (4)

f. $(S_C R_L S_{Sh}{}^-) \gg (S_C R_R S_{Sh}{}^-)$ By (4) and (5), inference and incomplete generalization.

g. $V_{S(L)} \ll V_{S(R)}$ By (6), both through extinction of the linkage $S_L S_F{}^*$ and establishing that of $S_L S_{Sh}{}^-$.

III. On the test run, considering the resulting potentials,

h. $_SE_{R(L)} = (V_{S(L)})(S_C R_L S_L) \ll (V_{S(R)})(S_C R_R S_R) = {_SE_{R(R)}}$ By (12) with (g).

i. $_SE_{R(L)} = (V_{S(F)}{}^*)(S_C R_L S_F{}^*) < (V_{S(F)}{}^*)(S_C R_R S_F{}^*) = {_SE_{R(R)}}$ By (12) with (e).

j. $_SE_{R(L)} = (V_{S(Sh)}{}^-)(S_C R_L S_S{}^-) < (V_{S(Sh)}{}^-)(S_C R_R S_{Sh}{}^-) = {_SE_{R(R)}}$ By (12) with (f).

Now whatever may be the laws of summation for reaction potentials sharing elicitor and R term, so long as it is some increasing function of the components, it is clear that

$$_SE_{R(L)} < {_SE_{R(R)}}$$

which is the desired result.

Consider as a final example the Spence-Lippitt type of latent-learning design reported as positive by those authors in 1940 (13) and by us in 1947 (8). Rats are run in a satiated state to goal boxes which, however, contain food or water. On the test run a state of either hunger or thirst is induced, and the appropriate choice tends to occur. Suppose food is on the right, water on the left. Then the sequences which occur during the satiated phase are

$$S_C \rightarrow R_R \rightarrow S_R S_F{}^*$$
$$S_C \rightarrow R_L \rightarrow S_L S_W{}^*$$

which will mnemonize two expect-

ancies $(S_CR_RS_RS_F^*)$ and $(S_CR_LS_LS_W^*)$ provided that "removal" is reinforcing (using [6] and possibly [7]). An additional contribution to the growth parameter might be made if we interpret (9) to mean a function of the form $V = a + b\,(D)(C^*)$ rather than of the form $V = b(D)(C^*)$, so that some valence persists for a cathected stimulus even when the need or drive variable is effectively zero. This is the approximate analogue in an expectancy theory to the claim in S-R-reinforcement theory that a stimulus retains its secondary reinforcing properties when the primary drive is satiated. That question, formulable in either system, is presently awaiting clear-cut solution experimentally (cf. **3, 4, 7, 10, 11**).

So after N trials of about equal runs to both sides,

$$(S_CR_LS_LS_W^*) \doteq (S_CR_RS_RS_F^*) \quad \text{by (1).}$$

And, since

$$(V_{S(W)}^*) \doteq (V_{S(F)}^*)$$

(assuming approximately equal pre-experimental experience with food and water and current near satiation for each) we have by the activation postulate

$$\begin{aligned} {}_SE_{R(L)} &= (V_{S(L)S(W)}^*)(S_CR_LS_LS_W^*) \\ &\doteq (V_{S(R)S(F)}^*)(S_CR_RS_RS_F^*) \\ &= {}_SE_{R(R)}. \end{aligned}$$

We now manipulate drive, making the animal very hungry but not thirsty. Then

$$V_{S(W)}^* \ll V_{S(F)}^* \quad \text{By (9)}$$

since

$$\begin{aligned} V_{S(W)}^* &= (D_W)(C_W)^* \\ &\ll (D_F)(C_F)^* = V_{S(F)}^* \end{aligned}$$

assuming from past history and (11) that

$$C_W^* \doteq C_F^*.$$

Then, by the activation postulate,

$$\begin{aligned} {}_SE_{R(L)} &= (V_{S(L)S(W)}^*)(S_CR_LS_LS_W) \\ &< (V_{S(R)S(F)}^*)(S_CR_RS_RS_F) \\ &= {}_SE_{R(R)}. \end{aligned}$$

There are certain additional consequences of the derivation that are of interest. To mention some of them briefly:

1. For a given valence the difference in ${}_SE_R$ is a function of the magnitude of the expectancies, since

$$\begin{aligned} &{}_SE_{R(L)} - {}_SE_{R(R)} \\ &\qquad = (V_L - V_R)(S_CRS) \end{aligned}$$

where

$$S_CR_RS_L = S_CR_RS_R = S_CRS.$$

a. But the magnitude of the expectancy increases with occurrences of the sequence, so that the percentage of appropriate responses on the test run should be less for rats who have had fewer "latent" runs during the satiated phase.

b. The growth rate of each expectancy depends on the valence of its expectandum (Mnemonization Law). For a constant number of cognition runs, the expectancy levels reached should be raised by any procedure that raises the valence of the goal boxes, even though *equally*. Thus, if another rat or home cage or preferred illumination is found in *both* boxes during the latent phase, this should raise the percentage correct on the test run.

However, plotting percentage "correct" on the test run against *number* of runs for groups differing in the valence of goal boxes should yield a family of curves of different slope but all approaching the same asymptote.

c. If the two goal boxes are varied from run to run so that the "modal" goal box on a side occurs randomly with a frequency P (and the presentation of food and water occur consist-

ently in this box) the percentage of correct choices on the test run is a function of P. The members of the family of curves in this case should approach different asymptotes.

2. For a given magnitude of expectancy the $_sE_R$ difference is a function of the valence difference. If both valences are small, $_sE_{R_R} - {}_sE_{R_L} = (V_R - V_L)(S_1RS_2)$ will be smaller than if one valence is large and the other small.

By the Law of Cathexis (11), both the cathexis of food and water will de minimal for rats that have never been on deprivation schedules for these commodities. As a result of this, by the Law of Mnemonization (1), expectancies for food on one side and water on the other will grow very slowly for such "naive" rats, and a short series of latent runs will, therefore, leave them with weak expectancies. Finally, by the Law of Valence (9), the valences of these expectanda will be minimal for such rats; and this is true both during the acquisition and test phases. Therefore, the $_sE_R$ difference between R_L and R_R on the test trial is doubly reduced by minimizing both the expectancy and valence terms in each (cf. Christie, **1**, **2**).

3. On the other hand, the *expectancy* strength approaches the same asymptote although more slowly, even for low valences (mnemonization law). If repeated occurrences of the sequence $S_1 \rightarrow R \rightarrow S_2$ have brought (S_1RS_2) close to this asymptote, we can then (outside the maze!) induce a strong cathexis in the alimentary expectanda by use of a feeding cycle. Such rats should (Law of Valence, Law of Activation) show the same per cent correct on a test trial as rats who had been hunger experienced prior to the experiment. But the asymptote of the expectancy (S_1RS_2) must have been reached.

Thus, for small numbers of "cognizing" runs, an interpolated hunger cycle (after the training series but before the test runs) ought not to bring these rats to the level of the previously hunger experienced. But for larger numbers of cognition runs, it should. The percentage correct on test runs plotted as a function of the number of training runs should approach the same asymptote for the interpolated group. But at the early phases of these curves a large difference is expected.

4. Since part of the valence of both S_R and S_L is assumed to be nonalimentary, based on the reward of removal, removal to a less valenced locus, e.g., mazelike box instead of home cage, or to a slightly negative-valenced locus such as a too-bright white box, should yield a slow growth of both expectancies. So for a small number of satiated runs, each expectancy has a lower value at the test run. Hence we expect less $_sE_R$ difference and more errors.

The obvious Hullian complaint would be, "Here you really admit the importance of the reward *during* the latent (satiated) phase." In a sense this is true. Yet, how does the reward *act* in the present formulation? It is not conceived of as generating a "habit," in favor of the later-to-be-chosen side. Suppose we were to give a *strong* but *equal* reward on both sides. The present derivation assumes that if we consider only rats having no bias to the left during the latent phase, they will still show the advantage in question on the test run. After all, the quantitative results of other studies indicate a very marked effect on response strength when an alimentary drive is raised from near zero to a 24- or 48-hour value. One would presumably be safe in predicting that the parameters of our equations are such that a slight

right-going bias (say, 10–20 per cent) manifested during satiated runs will be easily overcome by the valence shift on the test trial. Hullian postulates would in such a case merely raise the multiplication factor $f(D)$ and the algebraic sign of $({}_sE_{R(L)}) - ({}_sE_{R(R)})$ would still be negative. Barring the use of r_g, the predicted Hullian result is for such "biased" rats to manifest the bias with higher probability on the test run.

A concluding clarification of our own position seems necessary. The preceding has been presented as one way of formulating the expectancy position. We consider ourselves chiefly identified with some form of S-R-reinforcement theory, and our sympathies remain with it. The foregoing development is, therefore, *not* presented as empirically confirmed or even as a strong contender. Actually, the question of its factual adequacy has played almost no part in our thinking, which was directed at explicating one current theory rather than at proposing an empirically supported system with maximal attention to the facts.

REFERENCES

1. CHRISTIE, R. Experimental naiveté and experiential naiveté. *Psychol. Bull.*, 1951, 48, 327–339.
2. CHRISTIE, R. The effect of some early experiences in the latent learning of adult rats. *J. exp. Psychol.*, 1952, 43, 281–288.
3. ESTES, W. K. Generalization of secondary reinforcement from the primary drive. *J. comp. physiol. Psychol.*, 1949, 42, 286–295.
4. ESTES, W. K. A study of motivating conditions necessary for secondary reinforcement. *J. exp. Psychol.*, 1949, 39, 306–310.
5. ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, 57, 94–107.
6. KENDLER, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, 36, 212–220.
7. MACCORQUODALE, K., & MEEHL, P. E. Cognitive learning in the absence of competition of incentives. *J. comp. physiol. Psychol.*, 1949, 42, 383–390.
8. MEEHL, P. E., & MACCORQUODALE, K. A further study of latent learning in the T-maze. *J. comp. physiol. Psychol.*, 1948, 41, 372–396.
9. MEEHL, P. E., & MACCORQUODALE, K. Some methodological comments concerning expectancy theory. *Psychol. Rev.*, 1951, 58, 230–233.
10. MYERS, J. The reinforcement value of the sight of food for rats that are not hungry. Unpublished M.A. Thesis, Univer. of Iowa, 1949.
11. SHEFFIELD, F. D., & ROBY, T. B. Reward value of a non-nutritive sweet taste. *J. comp. physiol. Psychol.*, 1950, 43, 471–481.
12. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
13. SPENCE, K. W., & LIPPITT, R. Latent learning of a simple maze problem with relevant needs satiated. *Psychol. Bull.*, 1940, 37, 429.
14. TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Century Co., 1932.
15. TOLMAN, E. C., & GLEITMAN, H. Studies in learning and motivation: I. Equal reinforcements in both end-boxes, followed by shock in one end-box. *J. exp. Psychol.*, 1949, 39, 810–819.

[MS. received December 27, 1951]

Psychological Review
Vol. 60, No. 1, 1953

THE MATHEMATICAL FORMULATION OF A UNIFIED FIELD THEORY

WILLARD E. CALDWELL[1]
The George Washington University

To what extent is it possible to incorporate such areas of psychology as classical conditioning, instrumental conditioning, perception, the Weber-Fechner laws, Helson's work on adaptation level, and certain aspects of the nervous system within a common mathematical frame of reference? Points relating to this problem will be offered through an analysis of the results of experiments and the discussions of Weber, Fechner, and Helson.

The author has stressed the fact in previous writings (**1**, **2**, **3**, **4**) that, under certain conditions, any stimulus might serve as a motivation and the reduction of that stimulus as a reinforcement to the organism. One of the experiments designed to test this point of view was performed by Caldwell and Mosman (**3**). In this experiment low temperature was utilized in a maze as motivation and high temperature in the goal box as reinforcement. Learning curves in terms of time and errors were established for albino mice. Another one of these experiments was performed by Caldwell and Sandler (**4**). This experiment consisted of using gaseous formaldehyde in a maze as motivation and its relative absence in the goal box as reinforcement. Learning curves were established for the albino mice utilized in this experiment.[2]

An advantage of a temperature-type experiment over the use of hunger and thirst as motivations is the possibility of measuring the stimulus in both its motivating and rewarding phases to a higher degree of accuracy. These experiments might be viewed from a psychophysical frame of reference where the concept of just noticeable difference is in a more dynamic setting than in terms of differences between stimuli. It might be thought of as differences between differences. The problem then may be stated as: *What is the increase in the stimulus in the maze minus the increase in the stimulus in the goal box necessary to produce a just noticeable difference in time and errors.*[3] It might be possible to find experimentally many of the differ-

[1] The author wishes to express his appreciation to John G. Tiedemann for editorial assistance in the preparation of this paper.

[2] A series of experiments utilizing different kinds of stimuli as motivation is being conducted by the author at the present time at The George Washington University comparative laboratory. The first of these consists of placing mice in a maze which is rotated about a central axis. When the animal reaches the center of the maze, the experimenter stops the rotation. The animal then remains in the center for ten minutes. The purpose here is to investigate the role of vestibular stimulation and its relative absence as motivation and reinforcement. The second of these experiments is similar to the one performed by Caldwell and Mosman except that the maze has a high temperature and the goal box has a low temperature. The third experiment consists of a maze which has a very bright light overhead and a very dark goal box. In all of these experiments control animals run the maze without the additional stimulus of temperature, light, or vestibular excitation. The ability to measure more accurately both qualitative and quantitative variations in the maze and in the goal box allows the experimenter the possibility of treating data from these experiments within a psychophysical framework.

[3] It is assumed that if the various combinations of temperature differences were tested, that different experimental and control animals would be used with each difference.

ences in temperature and construct a scale with the center being zero.

This same general approach might be applied to the design where formaldehyde was used as the stimulus, or to Jerome and Flynn's (13) experiment using light avoidance. It is this kind of experimental design which led the author to think of psychophysical phenomena in terms of differences between differences rather than differences between stimuli, or, to put it in other words, to think of stimulus as being a field with a difference in that field.

Helson (8, 9, 10) has been doing some very important theoretical and experimental work in the field of perception on the general problem of adaptation level and frames of reference. Helson (8, p. 2) defines adaptation level as follows: "For every excitation-response configuration there is assumed a stimulus which represents the pooled effect of all the stimuli and to which the organism may be said to be attuned or adapted." Helson deals with the problem of adaptation from many different angles. He emphasizes the effects of background upon judging stimuli. He also suggests many applications of this to social phenomena. Helson says:

> Thus if a weight is said to be heavy, or a light dim, or a painting beautiful it is because the object in question appears above or below the indifference point of functioning. Such points can be said to represent the adaptation-level of the organism with respect to given stimuli. We find, for example, that the neutral stimulus in a series of weights ranging from 200 to 400 grams is about 250 grams, while in a series ranging from 400 to 600 grams the neutral stimulus is about 475 grams. If a background or comparison stimulus of 900 grams is introduced, the neutral stimulus becomes 350 grams in the first case and 550 grams in the second case. The adaptation-level, in general, tends to be determined largely by the values of the series and background stimuli. The effects of past experience are usually not sufficient to displace the adaptation-level from within the stimulus continuum (10, p. 380).

Helson says:

> The concept of adaptation-level (AL) must not be restricted to the effects of prolonged adaptation to more or less constant stimulation with greatly reduced capacity for response as a final end-state. There is an AL for every moment of stimulation. It is a function of all the stimuli acting upon the organism at any given moment as well as in the past (8, p. 3).

Helson, in discussing the mathematical treatment of adaptation level, says:

> Choice of a logarithmic function is in keeping with the well-established Weber-Fechner law governing stimulus-response relationships over a fairly wide range as well as the law of diminishing returns operative in the wider field of social and economic behavior. The formula found adequate in vision proved to be:
>
> $$AL = K(A_0^3 \bar{A})^{\frac{1}{4}} \qquad (1)$$
>
> where K is a fractional constant, A_0 is the brightness of the background, and $\bar{A}$ is the logarithmic mean of the brightnesses of the samples on the background. From this formula it is seen that the AL is a weighted geometric mean in which background is loaded three times as heavily as the log mean of all samples in the field (8, p. 3).

Helson also discusses the relation between adaptation level and Weber's law wherein he says:

> If the judgment of a stimulus, X_i, depends upon its distance from adaptation-level, A, it seems reasonable to assume that the judgment is related to the number of just noticeable differences between A and X_i since "psychological" distance must be stated in psychophysical terms. We can derive psychological distance

from the physical magnitudes involved by dividing the stimulus distance, $X_i - A$, by the value of the j.n.d. in physical units also. According to Weber's law, the increment ΔX which must be added to obtain a just noticeable difference from a standard, X, is a constant fraction of X or:

$$\Delta X = kX, \tag{1}$$

where k is the Weber constant. According to our basic assumption that judgments are made with respect to A, we should use this value as the base on which the Weber constant k operates. However, since a particular stimulus is being judged at any given time, we need to take its value into account also and hence we take the average of the two as the basis for a j.n.d. from adaptation-level giving equation (2):

$$\Delta A = k\,(X_i + A)/2. \tag{2}$$

Equation (2) actually weights the stimulus immediately in perception and adaptation-level equally in determining the size of the j.n.d. and this, the simplest assumption regarding their relative importance, proves to be adequate for a wide variety of data (9, p. 302).

Weber dealt with stimuli in a somewhat static sense. His basic formula was:

$$\frac{\Delta st}{st} = K.$$

This formula stated verbally is: The just noticeable increase in a stimulus is a constant fraction of that stimulus. Fechner enlarged upon this conception and his psychophysical formula was:

$$S = C \log R.$$

This formula stated verbally means, sensations are proportional to the logarithms of their exciting stimuli.

An important theoretical question at this point is the possibility of constructing a mathematical formula which will integrate the work of Weber, Fechner, Helson, and the maze-learning type of experiments discussed earlier. Another problem would be the utilization of such a formula in bringing together the fields of perception, learning, and classical and instrumental conditioning under one quantitative frame of reference. It then seems logical to try to integrate present-day experimental findings in various fields with the historical work of Weber and Fechner.

In reviewing the temperature experiment previously discussed, the assumption is made that psychophysical phenomena operate within an adaptive framework. Possibly these temperature experiments reveal the adaptive nature of the process more clearly than do some other types of experiment. This same type of process could be involved in the weight-lifting experiments dealing with perceptual phenomena. For every weight that is lifted we might postulate from Helson's (10) experimental data on adaptation level, there will be a weight which weighs less and yet will elicit a judgment of equality from the subject. The physical difference in the above weights could correspond to the difference in temperature between the maze and the goal box. This frame of reference might also be applied to the classical conditioning type of experiments proposed by Pavlov and Liddell. In the classical conditioning technique, a neutral stimulus could be established for the conditioned stimulus which would be comparable to the neutral point in the temperature and weight experiments.

Discussion of a Field Concept of Stimulus

The concept of stimulus has had many different definitions and applications in psychology. It is the purpose of this paper to emphasize the dynamic field aspects of the concept of stimulus. It is suggested here that,

operationally, stimulus should be thought of in two phases: one, the measure of the application of the stimulus, and, two, the measure of the stimulus after a stated period of adaptation by the organism to it. It is suggested that one j.n.d. below a stimulus should represent the stimulus in terms of the amount of adaptation exhibited by the organism. The theoretical viewpoint taken of the stimulus is that it is a field of energy represented by a difference acting upon an organism. One part of the field is the initial measure of the stimulus and the other part is at least one j.n.d. below the initial stimulus.

A TENTATIVE MATHEMATICAL FORMULATION

The following is a presentation and discussion of a tentative mathematical formulation of the theory of adaptive differentiation to be used as a tentative guide to experimental research.

Weber's formula, once again, is:

$$\frac{\Delta st}{st} = K.$$

A tentative revision of Weber's formula would be:

$$\frac{\Delta Si - \Delta Sj}{Si - Sj} = K,$$

where (*a*) Si represents the initial stimulus, (*b*) Sj represents one j.n.d. below the initial stimulus. The verbalization of this formula is as follows: *The increment of difference in a stimulus field represented by the difference between the increment of the initial stimulus minus the increment of the stimulus which is one j.n.d. below the initial stimulus to produce a just noticeable differentiation in the organismic field of energy is a constant fraction of the difference in the stimulus field represented by the difference between the initial stimulus and the stimulus which is one j.n.d. below it.*

Once again, Fechner's formula is:

$$S = C \log R.$$

A tentative revision of the Fechner formula would be:

$$d = C \log (Si - Sj),$$

where (*a*) d represents organismic differentiations, (*b*) Si represents the initial stimulus, (*c*) Sj represents one j.n.d. below the initial stimulus, (*d*) C represents a constant. The verbalization of this formula is as follows: *Organismic differentiations equal a constant times the logarithm of the difference in the stimulus field of energy represented by the difference between the initial stimulus and a stimulus which is one j.n.d. below the initial stimulus.*

SOME TENTATIVE POSTULATES

The following are some postulates utilizing the above mathematical formulation.

I (A) *The increment of the difference between the stimulus in a maze and a stimulus in the goal box which is one j.n.d. below the stimulus in the maze to produce a just noticeable difference in time and errors is a constant fraction of the difference between the stimulus in the maze and the stimulus in the goal box which is one j.n.d. below the stimulus in the maze.*

(B) *Differentiations in maze situations in terms of time and errors are equal to a constant times the logarithm of the difference.*

The formula for (A) is:

$$\frac{\Delta MSi - \Delta GSj}{MSi - GSj} = K,$$

where (*a*) MSi represents the stimulus in the maze, (*b*) GSj represents the stimulus in the goal box which is one j.n.d. below the stimulus in the maze.

The formula for (B) is:

$$MD = C \log (MSi - GSj),$$

where (*a*) *MD* represents maze differentiations in terms of time and errors, (*b*) *MSi* represents the stimulus in the maze, (*c*) *GSj* represents the stimulus in the goal box which is one j.n.d. below the stimulus in the maze.

An operational definition of GSj. *GSj* is obtained by utilizing time and errors rather than sensation or judgments in psychophysical experiments. The problem is, *what is the least difference between the maze stimulus and the goal stimulus to produce a just noticeable difference in time and errors from that which would appear by chance or in a situation where there was no difference between the stimulus in the maze and the stimulus in the goal box.* In other words, there should be a statistically significant difference in time and errors between the experimental situation where there is a difference between maze and the goal box and the experimental situation where there is no difference between the maze and the goal stimulus.

II (A) *The increment of the difference between an initial conditioned stimulus and a conditioned stimulus which is one j.n.d. below it to produce a just noticeable differentiation (utilizing a conditioned response in animals as a substitute for judgment) of stimuli is a constant fraction of the difference between the initial conditioned stimulus and the conditioned stimulus which is one j.n.d. below the initial conditioned stimulus.*[4]

(B) *Differentiations of stimuli utilizing conditioned responses as judgments equal a constant times the logarithm of the difference between the initial conditioned stimulus and the conditioned stimulus which is one j.n.d. below it.*

The formula for (A) is:

$$\frac{\Delta CSi - \Delta CSj}{CSi - CSj} = K,$$

where (*a*) *CSi* represents the initial conditioned stimulus, (*b*) *CSj* represents the conditioned stimulus which is one j.n.d. below the initial stimulus.

The formula for (B) is:

$$CRD = C \log (CSi - CSj),$$

where (*a*) *CRD* represents differentiation of stimulus where conditioned responses are substituted for an act of judgment, (*b*) *CSi* represents the initial conditioned stimulus, (*c*) *CSj* represents the conditioned stimulus which is one j.n.d. below the initial conditioned stimulus.

III (A) *The increment of the difference between a stimulus and a stimulus which is one j.n.d. below the initial stimulus for the organism to elicit a just noticeable differentiation utilizing an instrumental conditioned response as the equivalent of judgment is a constant fraction of the difference between the initial stimulus and a stimulus which is one j.n.d. below it.*

(B) *Differentiations in terms of instrumental conditioned responses equal a constant times the logarithm of the difference between the initial conditioned stimulus and the conditioned stimulus which is one j.n.d. below the initial conditioned stimulus.*

The formula for (A) is:

$$\frac{\Delta Si - \Delta Sj}{Si - Sj} = K,$$

where (*a*) *Si* represents the initial instrumental conditioned stimulus, (*b*) *Sj* represents the instrumental conditioned stimulus which is one j.n.d. below the initial stimulus.

The formula for (B) is:

$$ICRD = C \log (Si - Sj),$$

[4] This has reference to the experiments where animals are trained to discriminate between different tones, etc., using the conditioned response as the index of the difference.

where (a) $ICRD$ represents instrumental response differentiations, (b) Si represents the initial instrumental conditioned stimulus, (c) Sj represents the stimulus which is one j.n.d. below the initial stimulus, (d) C represents a constant.

IV (A) *The increase in a difference between a stimulus and a stimulus which is one j.n.d. below the initial stimulus to produce a just noticeable differentiation of lifted weights is a constant fraction of the difference between the stimulus and the stimulus which is one j.n.d. below the initial stimulus.*

(B) *Differentiations of lifted weights equal a constant times the logarithm of the difference between the initial stimulus and the stimulus which is one j.n.d. below the initial stimulus.*

The formula for (A) is:

$$\frac{\Delta Si - \Delta Sj}{Si - Sj} = K,$$

where (a) Si represents the initial stimulus, (b) Sj represents the stimulus which is one j.n.d. below the initial stimulus.

The formula for (B) is:

$$DW = C \log (Si - Sj),$$

where (a) DW represents the differentiations of lifted weights, (b) Si represents the initial stimulus, (c) Sj represents one j.n.d. below the initial stimulus, (d) C represents a constant.

V (A) *The increment of the difference between the intensity of a stimulus of constant duration and the intensity of the stimulus of constant duration one j.n.d. below the initial intensity of the stimulus to produce a just noticeable difference in the frequency of nerve impulses within a single nerve fiber is a constant fraction of the difference between the initial intensity of the stimulus of constant duration and the intensity of the stimulus one j.n.d. below the initial intensity of the stimulus.*

(B) *The frequency of nerve impulses is equal to a constant times the logarithm of the difference between the initial intensity of a stimulus of constant time interval and the intensity of the stimulus which is one j.n.d. below the initial stimulus intensity.*

The formula for (A) is:

$$\frac{\Delta I_1 - \Delta I_2}{I_1 - I_2} = K,$$

where (a) I_1 represents the initial intensity of a stimulus of constant time interval, (b) I_2 represents the intensity of the stimulus which is one j.n.d. below the initial stimulus.

The formula for (B) is:

$$FNI = C \log (I_1 - I_2),$$

where (a) FNI represents the frequency of nerve impulses, (b) I_1 represents the initial intensity of the stimulus of constant time interval, (c) I_2 represents the intensity of the stimulus which is one j.n.d. below the initial stimulus, (d) C represents a constant.

VI (A) *The increment of difference between the time interval of a stimulus of constant intensity and the time interval one j.n.d. below the initial time interval of the stimulus to produce a just noticeable difference in the frequency of impulses within a single nerve fiber is a constant fraction of the difference between the initial time interval of the stimulus of constant intensity and the time interval one j.n.d. below the initial time interval of the stimulus.*

(B) *The frequency of nerve impulses equals a constant times the logarithm of the difference between the initial time interval of a stimulus of constant intensity and the stimulus which is one j.n.d. below the initial time interval of the stimulus of constant intensity.*

The formula for (A) is:

$$\frac{\Delta T_1 - \Delta T_2}{T_1 - T_2} = K,$$

where (*a*) T_1 represents the initial time of a stimulus of constant intensity, (*b*) T_2 represents the time interval which is one j.n.d. below the time interval of the stimulus of constant intensity.

The formula for (B) is:

$$FNI = C \log (T_1 - T_2),$$

where (*a*) *FNI* represents the frequency of nerve impulses, (*b*) T_1 represents the initial time of a stimulus of constant intensity, (*c*) T_2 represents the time interval which is one j.n.d. below the time interval of the stimulus of constant intensity, (*d*) *C* represents a constant.

Some Applications of This Formulation to Existing Psychophysical Data

Most of the preceding postulates can be fitted to general psychophysical data. A simple example illustrating the application of previous data to this reinterpretation of the Weber-Fechner fraction can be taken from Garrett (**5**, pp. 334–335). In this experiment there are two stimuli, a standard and a variable comparison stimulus. These two stimuli might be lines, lights, tones, etc. The two stimuli have a value of 100 each. If the comparison stimulus is increased until a value of 110 is reached, it is judged to be just noticeably different from the standard. The Weber fraction is 1/10.

The following are some calculations of j.n.d.'s based upon this particular fraction:

1. 100
2. 110
3. 121
4. 133
5. 146.41
6. 161.05
7. 177.16
8. 194.87
9. 214.36
10. 235.79
11. 259.37

Now, if the formula

$$\frac{\Delta Si - \Delta Sj}{Si - Sj} = K$$

is applied between steps two and three or the figures, 110 and 121, the judgment in terms of this present theory would be from 1 to 2 or from 100 to 110 respectively. The substitution in this formula would be:

$$\frac{11 - 10}{110 - 100} = \frac{1}{10}.$$

From the standpoint of the organism, one may think of stimuli which are obtained by the Weber-Fechner fraction as ascending or descending in *pairs* and not as separate stimuli. The calculation now allows this reinterpretation to fit any data where the Weber-Fechner fraction has been found to hold.

In many applications of this mathematical formulation the supporting evidence will come from already existing psychophysical data and the problem is merely reinterpreting the data in terms of a more dynamic point of view.

For postulate II on classical conditioning, the experimental work of Hovland (**12**), where psychophysical methods were applied to the conditioned galvanic skin response, may be utilized. For postulate III on instrumental conditioning, reference may be made to Lashley (**14**). For postulate IV on perception, one example has been discussed and others include standard weight-lifting experiments, etc. For postulate V on the frequency of nerve impulses in relation to intensity of stimulation, reference may be made to the data from

Hodgkin (11) where the relationship between the intensity of a stimulus and the frequency of impulses was investigated in a single motor fiber from a crab. For postulate VI on the relationship between duration of the stimulus and its relation to the frequency of the impulse, reference should be made to the work of Hartline (6) and Hartline and Graham (7) where they studied the impulse discharge of single photoreceptors of the horseshoe crab. Postulate I on maze learning may be tested by further testings of temperature differences in the experiment performed by Caldwell and Mosman (3).

SUMMARY AND IMPLICATIONS

The previous postulates represent an attempted homogenization of somewhat diverse fields in psychology. It should be kept in mind that this formulation is a somewhat ideal one and would be subject to the same limitations and criticisms that have been applied to the Weber-Fechner fraction. Also, in actual applications of it, particularly outside of ideal laboratory conditions, the work of Helson on background stimuli may be important.

The question may be raised as to the value of such a mathematical formulation as presented herein. The following are some suggestions pertaining to its value. It allows:

1. greater semantic unification of psychological theory;
2. a more precise quantification of field theory;
3. a psychophysical treatment of motivation, learning, and reinforcement;
4. mathematical proof of the similarities between classical and instrumental conditioning;
5. a more universal application of Helson's concept of adaptation level;
6. for more specific, quantitatively measurable hypotheses for research;
7. a dynamic interpretation of the Weber-Fechner law;
8. the increasingly important concept of homeostasis to be treated quantitatively.

The previous formulations were presented with the idea of suggesting areas of research. This theory termed *adaptive differentiation* and its mathematical formulation were presented, not as something new, but rather as a new organization with the idea that it might further stimulate experimenters and theorists to focus on the problem of qualitative and quantitative unification.

REFERENCES

1. CALDWELL, W. E. Adaptive conditioning: A unified theory proposed for conditioning. *J. gen. Psychol.*, 1951, **78**, 3–37.
2. CALDWELL, W. E. The theory of adaptive differentiation. *J. Psychol.*, 1951, **31**, 105–119.
3. CALDWELL, W. E., & MOSMAN, K. F. The role of temperature change as reinforcement. *J. Psychol.*, 1951, **32**, 231–239.
4. CALDWELL, W. E., & SANDLER, H. M. The role of gaseous formaldehyde as reinforcement in maze learning in albino mice. *J. Psychol.*, 1952, **33**, 47–56.
5. GARRETT, H. E. *Great experiments in psychology*. New York: Appleton-Century-Crofts, 1951.
6. HARTLINE, H. K. Intensity and duration in the excitation of single photoreceptor units. *J. cell. comp. Physiol.*, 1934, **5**, 229–247.
7. HARTLINE, H. K., & GRAHAM, C. H. Nerve impulses from single receptors in the eye. *J. cell. comp. Physiol.*, 1932, **1**, 277–295.
8. HELSON, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
9. HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.

10. HELSON, H. *Theoretical foundations of psychology.* New York: D. Van Nostrand, 1951.
11. HODGKIN, A. L. The local electric changes associated with repetitive action in non-medullated axon. *J. Physiol.*, 1948, **107**, 165–181.
12. HOVLAND, C. I. The generalization of conditioned responses; I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.
13. JEROME, E. A., & FLYNN, J. P. A multiple choice box using light aversion as motivation. Naval Medical Research Institute, Bethesda, Md., 1950, Project NM 000 019.01.01.
14. LASHLEY, K. S. The mechanism of vision: I. A method for rapid analysis of pattern vision in the rat. *J. genet. Psychol.*, 1930, **37**, 453–460.

[MS. received for early publication July 11, 1952]

Psychological Review
Vol. 60, No. 1, 1953

ADDITIVE SCALES AND STATISTICS

C. J. BURKE

Indiana University

Psychological measurements do not possess the simple properties of the scales obtained for those basic dimensions of physics which have been designated as "fundamental magnitudes." The implications of this statement for quantitative psychology have been extensively studied and discussed with varying evaluations and recommendations. Frequently the recommendations have been such as to alter statistical practices, had they been followed.

Certain writers, notably Boring (2) and Stevens (6), have maintained that such statistical concepts as the sample mean and standard deviation presuppose, at the very least, a scale of equivalent units of some kind, thus casting doubt on the theoretical validity of extensive reliance on the *t* test, analysis of variance, and other statistical techniques widely used with psychological data. The resulting distrust of such widely used procedures has prompted Comrey (4) to seek their justification outside the strict limits of the traditional logic of measurement.

It is the purpose of the present paper to analyze this issue and to show that the use of the sample mean and standard deviation does no violence upon the data, whatever the properties of the measurement scale. Thus, the use of the usual statistical tests is limited only by the well-known statistical restrictions.

The argument to be given can be conducted from the axioms of probability and the axiomatic basis of measurement, but such detailed treatment would be merely pretentious, since the results which are necessary to establish the basic point are familiar to almost all psychologists.

The Nature of Measurement Scales

(The term "object" as used below should not be restricted to its usual meaning of "physical object"; rather it is to be interpreted with sufficient breadth to give the statements throughout this section meaning for psychological as well as physical measurement provided that the axioms can be satisfied.)

Objects which can be ordered on the basis of a pair of (physical, psychological, or other) relations are said to define a "dimension." For such objects there are two relations, objectual equality and objectual less-than-ness and the objects and relations satisfy the axioms of order reproduced by Comrey (4). Each object can be tagged with a number so that the numbers will satisfy a corresponding set of axioms. Thus there is a correspondence between the two systems:

(A) [Objects, objectual equality, objectual less-than-ness.]

(B) [Numbers, numerical equality, numerical less-than-ness.]

For some objects and relations, a further step is possible. An operation for combining the objects, "objectual addition," can be found such that the system:

(C) [Objects, objectual equality, objectual less-than-ness, objectual addition]

satisfies four additional axioms of combination (Comrey [4]). When this is the case the objects can be tagged with numbers so that the system:

(D) [Numbers, numerical equality, numerical less-than-ness, numerical addition]

satisfies four corresponding additional axioms of combination. Thus, in this case, there is a correspondence between (C) and (D).

When the systems (C) and (D) exist and correspond, we say that the objects define an "extensive dimension" and the numbers an "additive scale." In this case, of course, the systems (A) and (B) also exist and correspond.

When the systems (A) and (B) exist and correspond, but the systems (C) and (D) do not, we say that the objects define an "intensive dimension" and the numbers a "rank-order scale."

These matters are discussed in great detail by Campbell (3) and, more adequately for psychologists, by Bergmann and Spence (1). Pertinent information is presented in papers by Comrey (4) and Gulliksen (5). For our purposes, it is important to note only that the classification of a scale as additive depends upon the presence or absence of a certain correspondence, expressed in sets of axioms, between the numbers of the scale and the objects to which they refer—with, of course, appropriate ordering relations and combinative operations for each.

The Nature of Statistics

Statistical methods serve two major functions for psychologists.

a. They are used to summarize the salient features of individual sets of data.

b. They are used to test for differences between different experimental groups.

We shall discuss the second function in some detail, restricting our discussion to the simple case in which two groups are compared. In the typical psychological experiment the operations performed by the experimenter yield two or more sets of numbers. (In fact, unless the data exist in numerical form, means and standard deviations cannot be computed and the data are irrelevant for the present discussion.) It is obvious that two experimental groups will be judged alike or different in a given respect according as the collections of numbers classifying them in this respect are judged to be alike or different. It should be emphasized that we are here comparing the two sets of numbers *as numbers* and nothing else about them matters until after the statistical test has been made. The application of statistical techniques reflects merely our recognition of the unreliability of the small sets of numbers we have obtained and our unwillingness to perform the experiment again and again to determine whether the direction of the difference between our groups is reliable. We conceptualize a larger set of numbers, the statistical population, from which the sets of numbers we have obtained are two small samples, and inquire into the likelihood of two samples as disparate as we have observed arising from the given population. In answering this question, we often use the sample means and standard deviations as indices of important aspects of our collections of numbers. No interpretation other than this indicial one is intended. Means and standard deviations are used because they can always be computed, since numbers can always be added, subtracted, multiplied, and divided, and because means and standard deviations, conceived of merely as the results of operations with numbers, behave in certain lawful statistical ways.

In summary, the statistical technique begins and ends with the numbers and with statements about them. The psychological interpretation given to the experiment does take cognizance of the origin of the numbers but this is irrelevant for the statistical test as such.

Obviously, the same argument applies directly to the first function of statistics as well. The statement "The mean of these scores is 121" conveys in general

the same kind of information as the statement "The median of these scores is 122."

The objection that a well-established unit is necessary before the mean and standard deviation can be computed since their value is altered by a change in the absolute value of the scores (Comrey [4]) loses cogency when one notices that the mean and median will be affected in precisely the same way by adding a given number to each number in the sample and that the standard deviation and the interquartile range will be changed in the same way by multiplying each number in the sample by a given number.

AN EXAMPLE

To establish the point in another way, we consider an example of a statistical test based on an additive measure. Suppose that we are presented with two sticks, A and B, of apparently equal lengths, fixed on opposite sides of a room, and asked which is longer. We measure them and obtain a larger number for A. The two numbers, however, are nearly the same and we decide to repeat the measurement "just to make sure." On this occasion, we obtain a larger value for B. We proceed until we have 100 measurements on each stick and wish to answer the question without taking further measurements.

There are two collections of 100 numbers each, one for stick A and one for stick B. We test the hypothesis that they differ only through the unreliability of the measurements. A moment's reflection will show that we are not at all concerned with the additive nature of the scale for length. In adding the 100 numbers to obtain a mean for the measurements on stick A, we treat them as numbers and as nothing else. We make no interpretations whatever about adding 100 sticks together—there are only two sticks. Moreover, our interpretation by means of the t test is unaffected by the choice of length units we have made, provided that the units are the same for the two sticks.

It is seen that the comparison of the sets of measurements on the two sticks differs in no essential way from the comparison of two sets of IQ's.

SUMMARY AND CONCLUSIONS

We have noted that: (*a*) The properties of a scale of measurement involve correspondences between sets of axioms about objects and numbers, with appropriate relations and operations. (*b*) Statistical methods begin and end with numbers.

From (*a*) and (*b*), we have deduced that the properties of a set of numbers as a measurement scale should have no effect upon the choice of statistical techniques for representing and interpreting the numbers.

REFERENCES

1. BERGMANN, G., & SPENCE, K. W. Logic of psychophysical measurement. *Psychol. Rev.*, 1944, **51**, 1–24.
2. BORING, E. G. The logic of the normal law of error in mental measurement. *Amer. J. Psychol.*, 1920, **31**, 1–33.
3. CAMPBELL, N. R. *Physics, the elements.* London: Cambridge Univer. Press, 1920.
4. COMREY, A. L. An operational approach to some problems in psychological measurement. *Psychol. Rev.*, 1950, **57**, 217–228.
5. GULLIKSEN, H. Paired comparisons and the logic of measurement. *Psychol. Rev.*, 1946, **53**, 199–213.
6. STEVENS, S. S. On the theory of scales of measurement. *Science*, 1946, **103**, 677–680.

[MS. received January 12, 1952]

Psychological Review
Vol. 60, No. 1, 1953

THE AMES OSCILLATORY EFFECT: A REPLY TO PASTORE

FRANKLIN P. KILPATRICK

Princeton University

Pastore's (3) critical discussion in this journal of a portion of Ames's (1) monograph on the rotating trapezoid has prompted this rejoinder based on close association with Ames during the time he was preparing his monograph and on a number of my experiments in which the rotating trapezoid was employed as basic apparatus (2). Discussion will be confined to Pastore's comments concerning the generality of the oscillatory phenomenon, and to his asserted disagreement with Ames's explanation.

With regard to the generality of the phenomenon, Pastore says, "Ames . . . implies that the perceived oscillatory motion is unique to the trapezoidal window. The writer [Pastore], however, found that the oscillatory effect could be obtained with a wide range of forms" (3, p. 319). Ames will no doubt be horrified that such an implication is to be found in his paper, but a careful reading of it would seem to mark him as at least partially guilty on that count. The implication is unintentional, however, as Ames experimented with a great variety of forms which, when rotated, yield some degree of apparent oscillation. A number of these forms are described in the appendix to his paper (1). Ames's final selection of a form was based on his desire to maximize the amount and stability of the apparent oscillation so the phenomenon could be studied more conveniently and so it would not break down easily when placed in conflict with other moving objects. In his concentration on these ends, Ames failed adequately to point out the generality of the phenomenon, and Pastore must be credited with having performed a most necessary service.

However, it should be noted that such generality covering such a very wide range of forms as that described by Pastore is probably an artifact of Pastore's method. Pastore says, "Although the writer used smaller figures than did Ames, it is most likely that similar results will be obtained if larger figures were employed" (3, p. 319–20). On the contrary, it is most unlikely that similar results will be obtained if observation distance remains the same. Pastore's assertion overlooks the vital roles played by changing horizontal and vertical visual angles in the perception of either oscillation or rotation. Let us suppose that a rectangular plane figure is rotated once completely, starting at a position normal to the line of sight (0°). The horizontal visual angle (or the horizontal dimension of the retinal pattern) is at its maximum at 0°, diminishes until it is at a minimum at 90°, when the figure is seen end-on, increases to its maximum at 180°, decreases to its minimum at 270°, and finally reaches its maximum again at 360°. Exactly the same changes take place if the plane is turned to 180° and then reversed to 0°; that is, if it oscillates. Additional "information" is needed, then, to distinguish rotation and oscillation. The possible sources of information are many, but in the case of objects with relatively homogeneous surfaces and balanced lighting (such as those used by Pastore), the most important source is the asymmetrical alterations in the vertical dimensions of the retinal image. These are not the same for rotation and oscillation, and in most cases serve very well in differentiating the two. However, to the degree that the figures are made small or ob-

servation distance is increased, these asymmetrical alterations in the vertical dimensions of the retinal image become less effective, and the motion of the figure becomes more ambiguous. It may be seen either as oscillating or as rotating.

Pastore gives us little data concerning figure size or observation distance, but if the 1½" × 5" figure and the 35-foot observation distance he gives in one instance may be taken as representative, it is almost certain that many of his results may be accounted for in terms of the above analysis. This is not the phenomenon described by Ames. The illusory oscillation he described is quite unequivocal with a large figure at a near distance; in fact, within limits, and for monocular observation, the larger the figure and the nearer the distance, the more unequivocal it becomes. Ames used in his laboratory work a figure 19½" long, 23⅝" high at one side, and 12½" high at the other, with a viewing distance of about 10 feet (monocular). With this apparatus most observers report that some of the illusory effects are still present even when they move forward to a distance of 5 or 6 feet. Clearly, some other explanation is called for, and that is what Ames endeavored to supply in his monograph.

It is with this explanation that Pastore says he disagrees. But does he disagree? Ames says,

> As the trapezoidal window starts to rotate . . . [from a position normal to the line of sight with the short edge to the left] . . . , the total horizontal angle that the trapezoidal window subtends to the eye decreases. At the beginning of this decrease the trapezoidal window appears tipped back on the left. It has been learned from past experience with rectangular forms that a decrease of the total horizontal angle of our retinal images of a rectangularly perceived form which appears tipped away from us on the left could only take place if the side on the left went farther away. If it came nearer, the total horizontal angle of our retinal images would have to increase. So we interpret this decrease in the total horizontal dimension of our retinal stimulus pattern as a going-away of the left side of the window. That is, the window appears to reverse its direction of rotation and as the left side of the window keeps coming toward us, it appears to be going farther away. A similar apparent reversal is seen to take place when the trapezoidal window has revolved to a position where the short side . . . is to the right (1, pp. 15–16).

The heart of Pastore's "alternative interpretation" which ". . . seems to be in closer accord with the facts and also enables one to predict the occurrence of the oscillation effect" (3, p. 321) is as follows:

> Let us now return to the original assumed position of the trapezoidal figure—viz., at a tilt of plus 45 degrees. It so happens that the usual *O* will perceive the tilt to be not plus 45 degrees but some negative acute angle. That is to say, the long edge of the figure will be seen as being closer than the shorter edge. Despite the fact that the perceived tilt does not coincide with the actual tilt, the retinal projection of the rotating trapezoidal figure must remain an expanding one. In the phenomenal experience of *O* this could only be the case when the perceived direction of rotation of the trapezoidal figure is counter-clockwise (3, p. 321).

Ames chose for his illustration the portion of the figure's rotation when the retinal image is contracting; Pastore chose the portion when it is expanding. Otherwise the explanations are identical, unless Pastore means by "the phenomenal experience of *O*" something other than the sort of experience specified by Ames.

It may be that he does mean something different, because in one instance (3, p. 320) he quotes a statement by Ames concerning the nature of the experience involved, and then asks, "But

how could the individual, in his actions undertaken in an environment in which the retinal impression of rectilinear [1] forms is most frequently nonrectilinear, form an assumption about rectilinearity which is claimed to be so effective in shaping the sensory material? Why not form an assumption of 'trapezoidalness'?" (3, pp. 320–21). These questions indicate a misunderstanding of Ames's point. Ames does not suggest that the crucial experience is with a retinal pattern; he does not imply that the observer is looking at his own retina. On the contrary, the crucial experience is that of manipulating, dealing with, acting toward rectangular forms. It seems likely that we would, indeed, form assumptions of trapezoidalness if we lived in a world of trapezoidal forms.

Perhaps the key to this misunderstanding lies in what appears to be a preference by Pastore (3, pp. 321, 323) for a purely Gestalt explanation divorced from empiricism. If so, the issue is a clear one. Ames would agree ". . . that the preferred perceptual tendency is the outcome of the distribution of central processes (which are functionally related to retinal impressions) . . ." (3, p. 321), but would argue that the distribution and functional relationship must be accounted for mainly in terms of past experience. If this is the true issue, however, one still wonders what Pastore means by "the phenomenal experience of *O*."

Another matter regarding Pastore's "alternative interpretation" should be mentioned. He says, "One point of the interpretation proposed by the writer [Pastore] which remains unexplained concerns the factors responsible for a perceived surface whose tilt differs from the actual surface" (3, p. 323), thus giving the impression that the point is not central to Ames's explanation and perhaps was not even considered by Ames. On the contrary, Ames explicitly raised this point (1, pp. 7, 13, 16) and dealt with it at length. In fact, he even went to the trouble of preparing a 49-figure chart (1, p. 6) whose major function is to illustrate his discussion of that particular problem.

Incidentally, Pastore errs in his description of the role played by the discrepancy between perceived and actual tilt. He states, "It will be shown that oscillation (or reversal of the actual rotation of the plane figure) occurs when *O*'s perception of the tilt of the plane does not coincide with the actual tilt" (3, p. 321). There are two possible interpretations of this assertion, (*a*) that perceived and actual tilt are disparate *only* during the reversal, or (*b*) that during the reversal perceived and actual tilt do not coincide. If he means the first, he is clearly wrong. It is true of some figures (a rectangle, for example), but the only time the trapezoid is seen in its true position is when it is end-on. Then there is no tilt to be seen, just a line. If he intends the second interpretation, the statement is tautological and has no explanatory value. *Of course* perceived and actual tilt do not coincide during the illusory reversal.

It seems fair to say, in summary, that although Pastore and Ames may be in basic disagreement, it is not yet clear wherein that disagreement lies. The supposedly contradictory evidence Pastore reports was secured under conditions not at all comparable to those which obtained in the experiment he criticizes, and a major share of his discussion is concerned with issues which a careful study of Ames's monograph shows to be nonexistent.

[1] Pastore no doubt means "rectangular" (the term used by Ames) rather than "rectilinear." A trapezoid is rectilinear (i.e., bounded by straight lines), but not rectangular.

REFERENCES

1. AMES, A., JR. Visual perception and the rotating trapezoidal window. *Psychol. Monogr.*, 1951, **65**, No 7 (Whole No. 324).

2. KILPATRICK, F. P. Assumptions and perception: Three experiments. In F. P. Kilpatrick (Ed.), *Human behavior from the transactional point of view.* Hanover, N. H.: Institute for Associated Research, 1952.

3. PASTORE, N. Some remarks on the Ames oscillatory effect. *Psychol. Rev.*, 1952, **59**, 319–323.

[MS. received for early publication September 10, 1952]

Vol. 60, No. 2 March, 1953

THE PSYCHOLOGICAL REVIEW

FACTORS DECISIVE FOR RESUMPTION OF INTERRUPTED ACTIVITIES: THE QUESTION REOPENED[1]

MARY HENLE

Graduate Faculty of Political and Social Science, New School for Social Research

AND GERTRUD AULL

Wagner College

The question of the factors upon which resumption of interrupted activities depends has received surprisingly little attention despite the importance of the interruption technique for the experimental study of human motivation. Two possible explanations of the resumption of unfinished tasks have been proposed by Ovsiankina (5, p. 340), further investigated by Adler and Kounin (1), and subsequently discussed by Lewin (3, p. 820). These hypotheses define the problem with which the present paper is concerned:

1. The decisive factor for resumption might be the need tension created when the subject originally undertook to perform the task. The incompleteness of the task itself, then, would only be significant because the corresponding need remained unsatisfied. Resumption in this case would have the function of resolving the underlying need tension.

2. Resumption might arise primarily from the subject's perception of the incompleteness of the task itself and his recognition that the task requires action. This perception might be assumed to arouse a need for completion in the person. But the function of the resumption would be to meet the demands of the task, not primarily—or, at least, not originally—to satisfy the needs of the individual.

Both hypotheses, it will be noted, assign a role in resumption to the subject's need and to the incompleteness of the task, which possesses a valence for completion. But the first alternative makes the valence of the unfinished work a mere symptom of the already existing need tension, whereas the second hypothesis states that the forces responsible for resumption may issue primarily from the perception of the unfinished task itself. The two hypotheses are, of course, not mutually exclusive.

It might be argued that since we must assume that the needs of the person are involved in both cases, no real theoretical issue exists. But this does not solve the problem: the problem merely repeats itself on a different level. For then we must distinguish between the ego's needs in the service of a situation and those in the service of the individual (cf. 2).

There seems to be ample evidence that an unsatisfied need of the subject may lead to the resumption of an interrupted activity. Ovsiankina's result, a very high percentage of resumption of interrupted activities, was obtained with tasks which were trivial in nature, which

[1] This paper was prepared by Mary Henle when she was a Fellow of the John Simon Guggenheim Memorial Foundation.

did not urgently demand action from the subject. It was obtained, furthermore, with a variety of tasks, whose specific demands varied. There would appear, then, to be little doubt that a need tension within the subject may be decisive for resumption.

On the other hand, the evidence now to be reviewed would seem to exclude the second hypothesis.

Ovsiankina reports two sets of observations which lead her to reject or minimize the perceived incompleteness of the task as a factor decisive for its resumption.

1. She studied the part which the actual sight of the incomplete task played in 230 cases of resumption. In only a minority of cases (22 per cent) was resumption clearly related to the sight of the unfinished task. In a larger number of cases (33.5 per cent) the previous sight of the task was not necessary for resumption, as evidenced by subjects' introspections; also by instances in which the incomplete task was purposely hidden from the subject, who then had to ask for it or search for it in order to resume it; and by cases in which the subject, while in another room, announced his intention of resuming the unfinished work. The remaining instances were transitional cases in which, Ovsiankina states, it was difficult to decide just how much the sight of the task had to do with the resumption.

2. Ovsiankina undertook in another manner to decide whether the needs of the subject or the demands ot the task are decisive for the resumption of unfinished activities. Preliminary experiments were performed to discover whether resumptions would occur with unfinished tasks which had not originally been started by the subject himself and for which, therefore, no need tensions could be assumed.

Two tasks which had been used successfully in previous experiments were employed. They were left partially completed on a laboratory table. Only 3 of 11 subjects (27 per cent) resumed the unfinished work, although the same tasks showed resumption in 79 per cent of the cases in the main experiments with the interruption technique. By contrast with the prompt resumptions in the main experiments, one of the three resumptions occurred only after 40 minutes in the experimental room.

A repetition of the experiment with a different, more familiar task gave substantially the same results; only 1 of 11 subjects (9 per cent) resumed the incomplete task which was not his own, whereas the same activity had been resumed in 82 per cent of the cases in experiments with the interruption technique.

From the above evidence Ovsiankina comes to the following conclusions (5, pp. 345–6):

> A tendency to completion may issue from an incomplete activity. In our experiments, however, this tendency is extraordinarily weak by comparison with the forces driving to resumption which issue from the quasi-need arising from the interruption of work which has been begun. *The decisive factor, therefore, for the tendency to resumption in our experiments is the existence of a corresponding inner tension system.*
>
> Such an inner tension does not arise from the subject's perception of an incomplete task which is not his own, but presupposes the establishment, by instruction or occupation with the activity, of a special relationship between the subject and the work such that the person makes the work *his own.*

The last described experiment of Ovsiankina has been criticized by Adler and Kounin (1) in a number of respects:

1. The number of subjects employed was insufficient.

2. No attempt was made to ascertain that the incomplete task started by an-

other person was actually psychologically present in the subject's life space.

3. In the absence of specific instructions, the subject had no reason to believe that completion of the task was not prohibited.

Adler and Kounin devised experiments which are free of these difficulties and which they consider, therefore, to be more suitable for deciding whether personal need or perception of the incompleteness of the task is decisive for the resumption of interrupted activities. Their experiments were intended to create a situation in which two identical unfinished tasks were present in the subject's life space at equal distances from him. One of the tasks had been started by the subject himself ("I-task" or interrupted task), while the other was presented in a condition of incompleteness ("U-task" or unfinished task). Thus a quasi-need could be assumed only for the former. Should the "own" task be chosen for completion, Adler and Kounin state, this preference must be ascribed to the effect of the quasi-need within the person. If, on the other hand, no marked difference should be found in the amount of resumption of the two tasks then, Adler and Kounin propose, the need within the person must be eliminated as a decisive factor, and resumption would have to be accounted for by reference to the perception of the incomplete task alone.

The experiment of Adler and Kounin was performed with 22 children, ranging in age from about four to five years. Twenty subjects or 91 per cent resumed the own task (I-task), one subject (4 per cent) resumed the task started by another person (U-task), and another subject failed to resume either of the tasks. The resumptions occurred promptly and were generally followed by completion of the task.

According to the authors (1, p. 263) these findings

all argue for the same conclusion—resumption of the interrupted task depends here upon the established tension system corresponding to the quasi-need. The resumption of the U-task in only one case out of twenty-two argues against an interpretation of this resumption based upon the nature of the unfinished task, per se. Generalizing from these data we may say that the tension t(G) is a determining factor in resumption.[2]

These conclusions are in essential agreement with those of Ovsiankina cited above.

Lewin (3, p. 820) cites the work which has been described here as evidence that "the presence of uncompleted work of another person does not lead (or extremely seldom leads) to spontaneous completion in adults (Ovsiankina, 1928) or in children (Adler and Kounin, 1939). Both results indicate that the state of the need of the child is decisive for resumption."

While there are undoubtedly many cases which fit this description, a number of considerations make us question the general validity of these conclusions:

1. It is of interest to note that the conclusions of Ovsiankina, Adler and Kounin, and Lewin which have been cited here are not consistent with the general theoretical framework from which the studies in question evolved. Behavior, Lewin states, is a function of both the person and the psychological environment. He regards behavior as truly the result of an interaction between person and environment. But interaction cannot be regarded as a one-way affair, the person simply acting on a passive environment. We should therefore expect to find cases in which the forces underlying behavior issue primarily from the perceived object in the environment, and not only cases in

[2] "t(G) = the tension corresponding to some quasi-need . . ." (1, p. 257).

which the forces arise primarily in the person.

2. The conclusions drawn by Lewin, Ovsiankina, and Adler and Kounin are in conflict with certain evidence from real life situations. We know that there are situations which call for action and which are acted upon by observers who have had no previous part in the event. For example, we right a chair which is about to fall or prevent an accident by last-minute action. There are also cases in which people do resume the unfinished tasks of others, for example resuming and carrying forward research which leaves essential issues undecided, taking up and following through the ideas and ideals of others. The same applies to many situations which involve helping and replacing others.

3. Experimental evidence which appears in certain respects to contradict the conclusions of Adler and Kounin has been presented by Lewis (4). This investigator has demonstrated that in a truly cooperative situation the completion of a task by a co-worker may be just as satisfactory as completing the task oneself. More specifically, Lewis has shown that in a cooperative situation no difference exists in the amount of recall of self-completed tasks and interrupted tasks which the partner has completed. Zeigarnik (7), it will be recalled, demonstrated that unfinished tasks are favored in memory over completed ones. In Lewis' experiment, tasks left unfinished by the subject but completed by a cooperating partner behaved like completed tasks in recall.

Since these findings are opposed to the conclusions of Adler and Kounin, Lewis is led to call these conclusions into question. She writes (4, p. 206): "Actually, this experiment of Adler and Kounin missed the crucial point. Their results do not demonstrate that a person's needs may not directly conform to the objective requirements of the situation, but only that when the objective requirements are *identical*, a greater need arises to complete a task which one has personally begun."

Lewis' criticism may be warranted, but an even more fundamental difficulty has hitherto been overlooked. When comparing the effects of two conditions (here "ownership" vs. nonownership of a task), one assumes that the psychological situation is the same except for the condition being varied. The investigation of Adler and Kounin makes this assumption. These experimenters offered their subjects the choice between objectively identical I- and U-tasks for resumption. Their investigation rests on the assumption that these tasks were likewise *psychologically* identical for the subjects. This is something that must be established, however, rather than taken for granted. Once we consider the psychological situation, it is no longer self-evident that the two alternatives were alike in all respects except ownership. Before the argument of Adler and Kounin can be accepted, therefore, investigation is needed to determine the meaning of the I- and U-tasks for the subjects. In the absence of such investigation we are not justified in drawing conclusions about the factors responsible for the resumption of activities.

EXPERIMENTAL OBSERVATIONS

The first problem of investigation consists, therefore, in a re-examination of the work of Adler and Kounin. In order to do this, we report briefly a repetition of the experiment of these writers. The only important respect in which the present experiment differs from that of Adler and Kounin is that an effort was made to learn as much as possible about the subject's understanding of the experimental situation. Otherwise the experimental plan of the previous investigators was followed:

Two identical tasks, one started by the subject himself (I-task) and one left unfinished by someone else (U-task), were offered to the subject for resumption.

Experimental tasks. The tasks employed in this experiment were similar to those of Adler and Kounin. The I- and U-tasks included two identical colored cardboard houses, 10 inches in length by 6 inches in width. Before each house was a green wooden pegboard, 13 by 8 inches, each containing 192 holes. About 50 colored wooden pickets were provided with each house; these were to be fitted into the holes outlining the "yard" so as to construct a fence. In addition, some 30 colored pegs of various shapes, to be used as trees, bushes, flowers, etc. were supplied for each house and were to be used to make a garden. These materials were set out on a table, about 81 by 40 inches, in such a manner that the two tasks were about 40 inches apart.

The interruption task consisted of a small oilcloth duck partially stuffed with cotton wool. Additional stuffing was provided, and the child's task was to stuff the duck and pin it together.

Procedure. The materials for the I- and U-tasks were set out on the experimental table before the experiment began. The former (I-task) was left unstarted, all the pegs still contained in their boxes. The U-task had already been begun: about ten pickets had been placed into holes to start the fence, and one tree and two flowers had been set out in the garden. The *E* brought *S* into the experimental room and led him to the experimental table. The child was allowed a moment in which to survey the situation. Then *E* gave the following instructions:

> Look, Jimmy, what we have. Now I am going to tell you all about it. You see the two houses? (*E* leads *S* to the U-task.) This house has some trees and flowers in the garden and it has part of a fence around the yard. (*E* leads *S* to the I-task.) Now here is the other house. But it doesn't have a garden yet and no trees and flowers and no fence. And here in these boxes are all the things that you need for a fence and a garden. Now would you like to make a fence all around the house and make a pretty yard with all these trees and flowers?

When the I-task had been brought to approximately the same degree of completion as the U-task, *E* interrupted *S* by saying:

> Come over here, please, and see what I have here. Look at this nice duck. It's all flat; the cotton came out. I would like you to put some cotton in and make him nice and fat again.

When *S* had finished this task, *E* had intended to say: "Now, Jimmy, I have some writing to do. You go ahead and play by yourself with anything you want." These last instructions, however, never had to be used in this experiment.[3]

Careful records were kept of *S*'s behavior during the experiment and of all relevant spontaneous remarks. In addition, questions were occasionally asked by *E* in order to appraise the psychological situation of the subject.

Subjects. Twenty children, ten boys and ten girls, served as subjects in this experiment. They were pupils in a kindergarten class at the Gregory Avenue Public School in West Orange, N. J. They ranged in age from five to six years. Their IQ's ranged from 89 to

[3] Our experimental procedure, it should be noted, meets the conditions which Adler and Kounin consider to be necessary if the experiment is to provide a real test of the strength of the tendency to resume one's own task as compared with the work of another (1, p. 260): (*a*) To insure the psychological presence of both tasks for *S*, *E* introduced him to the U-task before allowing him to proceed with the I-task—the same method employed

140, with a mean of 107. *E* had become acquainted with the children during a two-hour session prior to the experiments.

Results. All 20 children resumed their own task after completing the interruption task. Our results thus offer striking confirmation of the quantitative findings of Adler and Kounin.

All resumptions occurred immediately and spontaneously; indeed, so prompt were the resumptions that *E* never had the opportunity to give the last part of the instructions. Noteworthy also was the marked reluctance of subjects to interrupt the I-task; repeated urging was generally needed. All *Ss*, furthermore, carried the I-task to completion.

The interpretation of Adler and Kounin, as stated above, rests on the assumption of the psychological identity of the I-task and the U-task. Specifically, it is assumed (*a*) that the two tasks are psychologically identical for the child, and (*b*) that they are psychologically on equal terms. Our qualitative data, now to be presented, challenge both these assumptions.

1. The qualitative evidence clearly shows that the two tasks are not psychologically identical. The child, in the course of working with the I-task, gives it a specific meaning or a new function which the U-task lacks. The I-task, in other words, comes to acquire new properties not possessed by the U-task. The demands of the two activities can thus no longer be considered to be alike. Two types of protocols may be cited as evidence for this difference.

a. A number of subjects made up stories around the I-task. Completion, then, no longer meant merely meeting the formal demands of the task as envisaged by *E* (demands also possessed by the U-task); the demands of the story, too, had to be satisfied. The following protocols are representative:

Subject 1: "I better hurry up or the dog will run out. . . . Their Great Dane will run out if I don't hurry." (Finishes fence.) "There, now you can't get out!"

Subject 4: "I'm going to make it *real* pretty. . . . This house is for sale. That's why it has to be pretty."

Subject 6: "I am going to put a lot of flowers in. . . . What are we going to do? The children will pick all the nice flowers. The mother just will have to put another fence around so they can't pick them. Just wait, children, in half an hour I will be done. Let's hurry, or they will pick them on the other side. . . . So, now they can't pick the flowers."

In these cases, then, the task is no longer the task as it was set by *E*. It is not just a matter of building a fence around a house; rather, the Great Dane must be prevented from running away or the flowers must be protected against children, or else a house must be made suitable for sale. These specific meanings and requirements are not possessed by the U-task. The I-task is thus not psychologically equivalent to the U-task. Its properties and requirements are different and, one also suspects, more compelling than those of the U-task.

b. Other *Ss*, prior to interruption, announced a plan of action with regard to the I-task.[4] Completion of the latter

by Adler and Kounin. (*b*) The final instructions, modeled after those of the previous investigators, were intended to make *S* feel perfectly free to resume either task after interruption or not to resume at all. Striking evidence of the freedom of the experimental situation is the fact that these additional instructions were never needed. (*c*) That the tasks possessed intrinsic appeal for our *Ss* was evidenced throughout by their behavior.

[4] This possibility has been mentioned by Torrey (6, p. 203). Commenting on the interpretation of Adler and Kounin, she writes: "The perception of a task which is merely seen partly done by another person and that of a task which the individual has himself worked on are themselves sufficiently different to warrant the prediction that they will in-

thus means the specific completion demanded by the plan. Again the properties of the I-task have changed as compared with those of the U-task. The following examples may be cited:

Subject 11: "I am going to use pretty colors and put trees right across the front."

Subject 19: "I want this tree right here; I like it. There is nothing going to be here, except all little flowers. My aunt has a house like this."

2. In a considerable number of cases, the qualitative evidence reveals, the I- and U-tasks come to be distinguished by the different roles which they play in the experimental situation. The U-task is perceived as a model, while the I-task becomes the main task.

a. Several *S*s expressly stated their belief that the U-task was introduced as a model. For example:

Subject 16: "I am going to make it (fence) all around, better than this" (pointing at U-task).

Subject 18, after staring at the U-task in fascination, happily starts his I-task. Later, pointing to the U-task, he remarks: "I bet you this (U-task) is to show how to do it and this (I-task) is to do. Right?"

b. Other *S*s clearly indicated by their behavior during the experiment, though they did not expressly state it, that they considered and used the U-task as a model for the I-task. They studied the U-task before starting their own, walked over to it while they were doing the I-task, ran back and forth between the two tasks or looked over to the U-task before making new moves on the I-task. This was particularly true of *S*s who otherwise displayed insecurity in their behavior. The following examples are typical:

Subject 2, a frightened little girl, kept looking back and forth between the U-task and her own, anxiously using the former as a model.

Subject 10, after finishing the fence, walks over to the U-task, her hands full of pegs, and looks at it again before she starts to work on the garden.

Subject 17, though evidently anxious to start the I-task, runs over twice to inspect the U-task before he begins work on his own.

Thus, by assuming the role of model, the U-task becomes psychologically different from the I-task. As with our other cases, one suspects that here too the difference between the two activities favors the resumption of the I-task. It seems safe to assume that the forces issuing from the main task are stronger than those issuing from a mere model, which is hardly a task at all in its own right.

Almost all *S*s of this experiment can be placed in one or more of the categories listed above. In a few cases the child was not sufficiently communicative and his behavior not sufficiently revealing to indicate the specific meaning of the experimental tasks to him. There seems, however, to be no reason to assume that the psychological situation was essentially different for these uncommunicative children than for the others.

Since the procedure of the present experiment was patterned closely after that of Adler and Kounin, and since the quantitative results agree well with theirs, it may be assumed that the essential conditions of these investigators have been reproduced here. But qualita-

fluence behavior differently. . . . Perhaps the well-formulated plan of how the task shall be done is the most important additional aspect of the task actually begun *by* the subject. The tension when this plan, which is the product of the subject's own imagination, is broken into would be much greater than what might be inspired by mere inspection of someone else's half-done work."

The present experiment, which was completed before the publication of Torrey's paper, shows that this is a good hypothesis, though not sufficient to account for the results of Adler and Kounin.

tive study reveals that under these conditions the I-task and the U-task, despite their physical identity, are by no means psychologically identical.[5] Furthermore, the differences between them appear to be such as to favor the resumption of the I-task. We can therefore not conclude, as do Adler and Kounin, that this task is preferred to the U-task because it is the individual's own, not someone else's activity. It might also be that the other differences in the properties and roles of the two kinds of tasks are responsible for the difference in resumption between them.

The question of what factor is decisive for the resumption of interrupted activities—the tension system within the person or the perceived demands of the task—is thus reopened by the findings of the present experiment. We are again led to ask: Will it always be true that a person deals with his own task in a privileged way? We have preliminary evidence which suggests that this is not always the case, that there are conditions under which a child is just as likely to resume and carry to completion the unfinished activity of another person as to start a task of his own. In any case, it is clear from the present investigation that it is by no means established that the needs of the individual are the only factors decisive for resumption.

[5] It might be argued that our experiment has only shown that, in the course of the individual's relations with an object, changes in the latter occur. This argument does not, however, do away with our present problem: we are here concerned with testing the assumption of Adler and Kounin that the properties of the I-task and the U-task are identical, so that a preference for the former in resumption is to be attributed to the quasi-need of the subject. The error of these investigators consists precisely in ignoring the changes which occur in a task in the course of the individual's relations with it.

Since the relation between the subject's own need and his resumption of an interrupted activity seems to have been amply demonstrated (though not by Adler and Kounin), the first task of future investigation in this area would seem to be to create situations in which the perceived requirements of the task determine resumption. Such investigation would be of significance for an understanding of the nature of the relation between the individual and his psychological environment as well as for such more specific problems as the present one of the factors determining resumption or the question of cooperative action.

The methodological implications of this study are clear. In the present field of investigation quantitative findings in the absence of qualitative study of the individual's understanding of his situation are likely to be misleading.

REFERENCES

1. Adler, D. L., & Kounin, J. S. Some factors operating at the moment of resumption of interrupted tasks. *J. Psychol.*, 1939, 7, 255–267.
2. Asch, S. E. *Social psychology.* New York: Prentice-Hall, 1952.
3. Lewin, K. Behavior and development as a function of the total situation. In L. Carmichael (Ed.), *Manual of child psychology.* New York: Wiley, 1946.
4. Lewis, Helen B., & Franklin, Muriel. An experimental study of the role of the ego in work. *J. exp. Psychol.*, 1944, 34, 195–215.
5. Ovsiankina, Maria. Die Wiederaufnahme unterbrochener Handlungen. *Psychol. Forsch.*, 1928, 11, 302–379.
6. Torrey, Jane W. Task completion as a function of organizational factors. *J. Pers.*, 1949, 18, 192–205.
7. Zeigarnik, B. Das Behalten erledigter und unerledigter Handlungen. *Psychol. Forsch.*, 1927, 9, 1–85.

[MS. received February 1, 1952]

Psychological Review
Vol. 60, No. 2, 1953

THE MEANING–FAMILIARITY RELATIONSHIP [1]

CLYDE E. NOBLE

Perceptual and Motor Skills Laboratory,
Human Resources Research Center

The present series of investigations represents an attempt to identify and quantify some of the attributes of verbal material relevant to human learning. The first paper of this series (10) reported an analysis of stimulus meaning (m). Evidence reviewed in McGeoch (7), Underwood (17), and Woodworth (18) indicates that the *familiarity* of a stimulus or a response item may also be an important variable operating in verbal learning. Some psychologists, however, have questioned whether meaning and familiarity are distinct concepts. Underwood (17), for example, suggests that "meaningfulness . . . is the same characteristic as implied by the term *familiarity*" (p. 411). Others (5, 9, 18) have indicated only that some positive correlation may obtain between these two variables.

The objectives of the present study are (a) an analysis of the attribute of familiarity in verbal stimulus material, and (b) an investigation of the functional relationship between meaning and familiarity.

An Analysis of Familiarity

Theoretically, the familiarity of a stimulus (or response) may be regarded as some function of its frequency of occurrence in an organism's history. Humans may experience familiarity with a verbal stimulus in a number of ways: through the special senses of vision or audition, by means of speech (spoken or implicit), or by writing. By these means familiarity may become a learned stimulus attribute. The learning process is at least descriptively clear. Frequency of stimulation—including that which is response-produced, or proprioceptive—is a necessary condition of learning. That it is not also a sufficient condition is attested to by the research reviewed by McGeoch (7). For verbal learning, as in many other types, some kind of reinforcement or knowledge of results seems jointly necessary.

Stimulus frequency is a variable to which a judge may be motivated and set to respond differentially. The empirical procedure of eliciting such differential responses might be by requiring S to point to a list of stimuli singly to indicate their degrees of *familiarity* to him. Important work related to this concept has been done by Thorndike (14), Thorndike and Lorge (15), and by Haagen (5). In brief, a consideration of the attribute of familiarity (f) reveals that it is a concept definable by the operations involved in a psychological (or psychometric) scaling method.

Assuming Thurstone's (16) law of comparative judgment, and utilizing the method of successive intervals developed by Saffir (12), Mosier (8), and Edwards (3), one may transform frequencies of rating into deviates of the normal curve to yield a set of scale values having the properties of an equal-interval scale. Hence, the familiarity of any stimulus, S, is given by

$$f_s = l + \left(\frac{.50 - cp}{p}\right) i, \qquad [1]$$

[1] The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or endorsement of the Department of the Air Force.

where l = lower limit of interval containing Mdn in x/σ units; cp = cumulative proportion below interval containing Mdn; p = proportion within interval containing Mdn; i = interval width on familiarity continuum in x/σ units.

Procedure

The same list of 96 dissyllables used in establishing the author's *m*-scale (10) was used to determine quantitative values for the attribute of stimulus familiarity. It was decided that a graphic rating schedule (method of single stimuli) was the most suitable defining task for this purpose. More elaborate methods, such as that of paired comparisons, would have been laborious and time consuming considering the number of stimuli involved. A rank-order method would have been inappropriate in view of the nature of the attribute investigated, since the act of ranking might have seriously altered the original values of the items. In addition, recent advances in the theory of measurement have provided a rationale and an internal consistency check for the present method. Finally, it was believed that the sample of 200 *S*s was adequate to insure a high reliability for the ratings.

The stimulus items were administered in mimeographed test booklet form. Twelve items appeared on each page, followed by spaces for graphic ratings arranged in columnar form. The complete list sequence was randomized with the aid of a table of random numbers, and the sequence and order of pages were counterbalanced by shuffling the attached answer sheets prior to administration. The sample totaled 200 airmen, selected from four different flights undergoing routine classification testing at the Human Resources Research Center. The *S*s were group-tested in four units of 50 men each, and the testing periods were held during August 1951. A test session was of approximately 20-min. duration. Instructions to the *S*s were as follows:

This is a test to find out how often you have come in contact with certain words.

You will be given a list of 96 nouns and you are to rate each one as to the number of times you have experienced it by simply placing a check mark (√) in one of the five spaces provided for your rating.

The five possible ratings are described by the words NEVER, RARELY, SOMETIMES, OFTEN, and VERY OFTEN. This means that you have *seen* or *heard* or *used* the particular word (in writing or in speech) either:

1. NEVER (You have *never* seen or heard or used the word in your life.)

or

2. RARELY (You have seen or heard or used the word at least once before, but only *rarely*.)

or

3. SOMETIMES (You have *sometimes* seen or heard or used the word, but not often.)

or

4. OFTEN (You have *often* seen or heard or used the word, but not very often.)

or

5. VERY OFTEN (You have seen or heard or used the word *nearly every day* of your life.)

Do not be bothered if you are unable to give a definition of some of the words. Simply rate each one as to the number of times you have come in contact with it regardless of its meaning.

There may be some words which you have *used* or *heard* more often than you have *seen* them. Or there may be other words which you have *seen* more often than you have *used* or *heard* them. In such cases, always give the word the *highest* rating of the three.

For example, you probably use or hear the word FLIGHT CHIEF often, but you may never have seen it in print. In this case, you would rate FLIGHT CHIEF as "Often," as shown on the line below:

FLIGHT CHIEF:

		Never	Rarely	Some-times	Often ✓	Very Often

The *S*s were then given four practice tasks, using as stimuli the words BREAKFAST, ZITHER, ENGLAND, and SYZYGY.

Instructions regarding motivation and set were as follows:

When the examiner gives you the "GO" signal you will turn the page and begin rating the list of words at your own speed. This is not a "speed" test, for each man will be given plenty of time to finish. The important thing is for you to be as accurate as possible.

Be as honest in your ratings as you can. Many of the words in this test are very rare, so you are not expected to have come in contact with all of them. Just make the best estimates you are capable of.

Pay no attention to those working near you, for the words are in mixed order and each man will be working on a different item.

Results and Discussion

The graphic method of recording *S*'s responses was simple, objective, and rapid. The index of familiarity (f) of a particular stimulus was defined in equation [1] as the median of the distribution of judgments on the familiarity continuum. Therefore, the method of successive intervals outlined by Edwards (3) was followed to compute the interval widths of the middle three categories. Failures of rating were not counted, the *n*'s of the corresponding arrays being decreased accordingly. Pro-

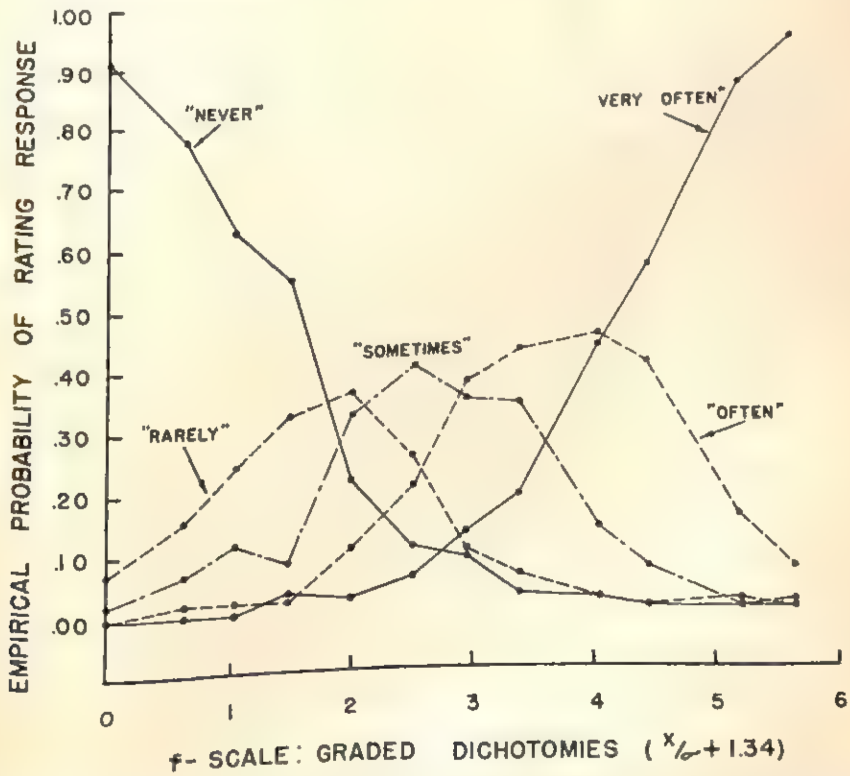


FIG. 1. Relative frequency distributions of graphic ratings in the five categories of the Familiarity Schedule for 12 representative stimulus words as functions of f-value derived from the method of graded dichotomies ($N = 200$)

TABLE I

LIST OF DISSYLLABLE WORDS (NOUNS) IN RANK ORDER OF INCREASING FAMILIARITY (f) AS DEFINED BY NORMAL DEVIATE TRANSFORMATIONS OF RATINGS ON FIVE-POINT GRAPHIC RATING SCHEDULE ($N = 200$)

Rank	Word	Word No.	f Value	Rank	Word	Word No.	f Value
1	POLEF	63	0.00	49	TANKARD	76	1.71
2	NOSTAW	55	0.10	50	NAPHTHA	51	1.73
3	GOKEM	25	0.16	51	TARTAN	78	1.85
4	LATUK	41	0.16	52	ROSTRUM	72	1.93
5	VOLVAP	86	0.17	53	ARGON	1	2.00
6	ZUMAP	96	0.19	54	YEOMAN	91	2.15
7	GOJEY	24	0.21	55	JITNEY	35	2.17
8	BYSSUS	8	0.39	56	VERTEX	84	2.17
9	XYLEM	90	0.40	57	RAMPART	68	2.21
10	SAGROLE	73	0.42	58	PERCEPT	61	2.30
11	TAROP	77	0.62	59	PALLOR	60	2.33
12	QUIPSON	66	0.64	60	PALLET	59	2.52
13	KUPOD	40	0.66	61	SEQUENCE	74	2.54
14	JETSAM	33	0.67	62	QUARRY	64	2.55
15	ULNA	81	0.67	63	ENTRANT	17	2.60
16	NEGLAN	53	0.69	64	MALLET	47	2.67
17	MEARDON	49	0.73	65	PIGMENT	62	2.67
18	BALAP	4	0.74	66	ZEBRA	93	2.74
19	DELPIN	13	0.76	67	TYPHOON	80	2.81
20	GAMIN	22	0.79	68	KENNEL	38	2.82
21	BRUGEN	7	0.85	69	ORDEAL	57	2.84
22	TUMBRIL	79	0.86	70	QUOTA	67	2.97
23	STOMA	75	0.89	71	ZENITH	94	2.98
24	BODKIN	6	0.90	72	FATIGUE	18	3.18
25	ICON	29	0.91	73	REGION	69	3.40
26	MATRIX	48	0.92	74	KEEPER	37	3.58
27	GRAPNEL	26	0.94	75	VILLAGE	85	3.67
28	NARES	52	0.96	76	GARMENT	23	3.69
29	KAYSEN	36	0.97	77	UNIT	83	3.70
30	LEMUR	43	1.01	78	JEWEL	34	3.71
31	FLOTSAM	21	1.02	79	YOUNGSTER	92	3.85
32	RENNET	70	1.06	80	ZERO	95	3.85
33	DAVIT	12	1.13	81	WAGON	87	3.93
34	MAELSTROM	46	1.13	82	EFFORT	15	3.96
35	ROMPIN	71	1.13	83	INSECT	31	3.97
36	WIDGEON	89	1.14	84	HUNGER	28	3.98
37	CAROM	11	1.18	85	INCOME	30	4.06
38	CAPSTAN	9	1.31	86	JELLY	32	4.11
39	OVUM	58	1.31	87	HEAVEN	27	4.19
40	FERRULE	20	1.32	88	CAPTAIN	10	4.27
41	ATTAR	3	1.33	89	UNCLE	82	4.34
42	NIMBUS	54	1.34	90	OFFICE	56	4.44
43	FEMUR	19	1.36	91	LEADER	42	4.47
44	LOZENGE	45	1.37	92	KITCHEN	39	4.71
45	WELKIN	88	1.49	93	QUARTER	65	4.76
46	BODICE	5	1.58	94	ARMY	2	4.79
47	ENDIVE	16	1.59	95	DINNER	14	5.22
48	LICHENS	44	1.67	96	MONEY	50	5.66

portions ≥ 0.98 and ≤ 0.02 were arbitrarily rejected, as is commonly done in the method of paired comparisons (4).

In view of the fact that for 51 of the 96 items more than 50 per cent of the judgments fell in either of the extreme categories, thereby constituting indeterminate successive intervals f-values, Attneave's (1) method of graded dichotomies was adopted to estimate the medians. These f-values are shown in Table 1, together with their reference numbers, ranging in rank order from dissyllables of low familiarity (e.g., rank 1: POLEF) to those of high familiarity (e.g., rank 96: MONEY). In order to orient the scale to an arbitrary origin, a constant of 1.34 was added to all values. The empirical range in x/σ units was then from 0.00 to 5.66.

One advantage of the method of successive intervals (or graded dichotomies) lies in the test of internal consistency which it affords. By using the $N + r - 2 = 99$ parameters one can reproduce the $N(r-1) = 384$ independent empirical proportions with an average error of 0.019.

Further evidence related to the internal characteristics of the f-scale is exhibited in Fig. 1. This graph plots the relative frequency distributions of the judgments in the five categories of the Familiarity Schedule for 12 stimulus words as functions of f-value. The 12 items were selected to represent approximately equal intervals along the scale. The regularity and symmetry of the probability functions is apparent.

Results very similar to these were obtained with a preliminary sample of 100 airmen in 1950. The rating schedule differed in that a category of DOUBTFUL was used instead of the RARELY category, but otherwise the stimuli and procedure were identical. The average

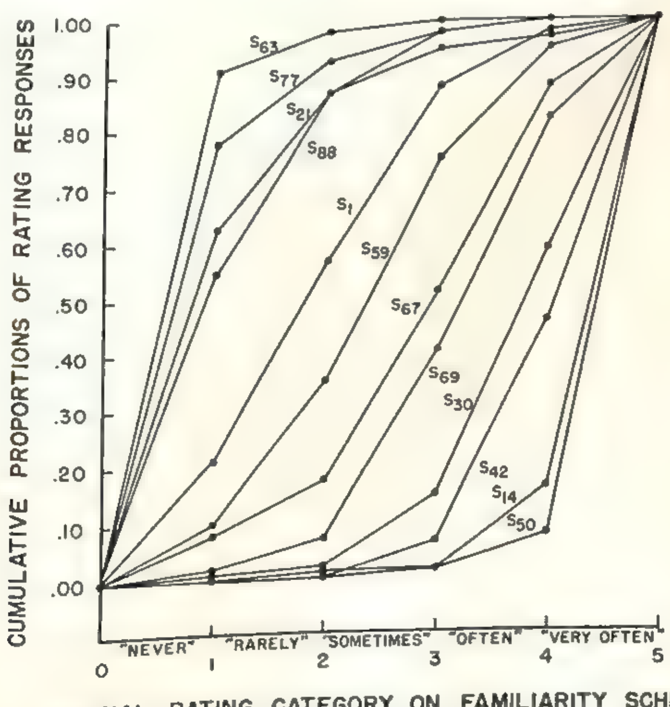


FIG. 2. Cumulative proportions of ratings in the five categories of the Familiarity Schedule for 12 representative stimulus words ($N = 200$)

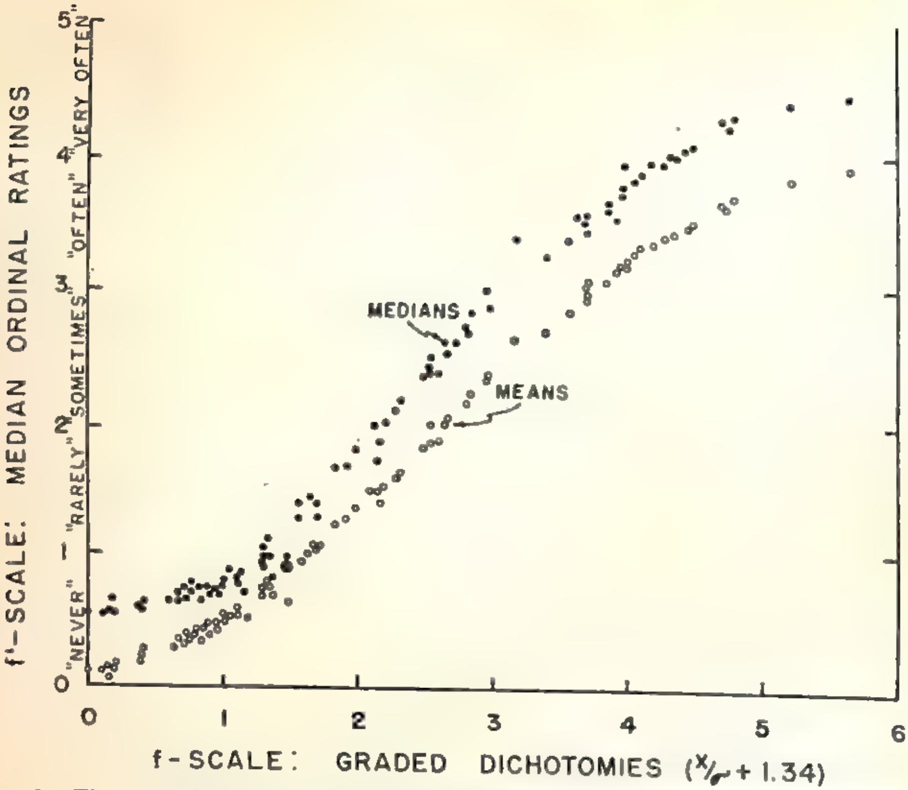


FIG. 3. The relationships among median and mean ordinal ratings and graded dichotomies scale values derived from the Familiarity Schedule ($N = 200$)

error was 0.023, and the probability distributions were similar to those in Fig. 1. This early form of the Familiarity Schedule was discarded because it was suspected that the DOUBTFUL category introduced a qualitatively different judgment into the task. There is now little evidence for this view ($r = 0.992$).

It is known that median scale values derived from the method of equal-appearing intervals are not linearly related to those treated by paired comparisons (6, 12), successive intervals (3, 12), or graded dichotomies (1, 3). The occurrence of "end-effects" is commonly attributed to the behavioral *inequality* of the rating intervals, such that skewed distributions of judgments arise at the extremes of the scale. The source of this nonlinearity is exhibited in Fig. 2. Here the skewness is seen to decrease from both ends toward the midpoint, where the cumulative distribution associated with stimulus No. 59 is quite symmetrical.

Figure 3 shows the consequent end-effects when either median ordinal ratings (f') or mean ordinal ratings (f'') are plotted against the graded dichotomies f-values. It should be recalled that the judges were not instructed to form equal-*appearing* intervals as is often done (4), hence the end-effects in this investigation are presumably uncomplicated by any artifact of instructions. The average category widths in x/σ units were: NEVER $= \frac{0}{0}$; RARELY $= 0.82$; SOMETIMES $= 0.90$; OFTEN $= 1.12$; and VERY OFTEN $= \frac{0}{0}$. Inspection of Fig. 3 indicates that, contrary to popular opinion (4), mean ratings are at-

tended by somewhat less distortion than are medians.

In order to satisfy the second objective of this study, the determination of the functional relationship between meaning and familiarity, a necessary condition is the possession of equal-interval scales for both variables. For the status of the m-scale, see (10). Since the logic of paired comparisons underlies the present scaling method, and since paired comparisons transformations are linearly related to those of successive intervals (3, 12) and to graded dichotomies (1, 3), then the unit-distance condition is satisfied for the f-scale.

THE MEANING-FAMILIARITY RELATIONSHIP

Several writers have called attention to the relationship between the frequency of appearance in writing and the number of synonyms of words in the English language. Zipf (19) and Thorndike (14) have shown this correlation to be a positive one.

With respect to another correlate of frequency—familiarity—certain psychologists hold that its relation to meaning is one of *identity*. Of course, the verification of this hypothesis is contingent upon variables which traditionally have been left undefined.

From logical considerations, the author prefers to regard any hypothetical meaning-familiarity relationship not as an identity, but rather as a determinate correlation between two different attributes. An *outré* example will serve to clarify the logic underlying this expectation. Consider an S–R connection which has been reinforced fifty times; that is to say, a stimulus element, S_i, and an associated response element, R_i, have thus occurred sequentially. Assume further that the sequence $S_i \rightarrow R_i$ is *unique*; i.e., that only S_i and R_i have ever been so connected. Clearly then, by comparing S_i with other stimuli S_j, S_k, to each of which several responses have become conditioned, its owner may exhibit S_i to be relatively high in f-value but quite low in m-value. The logical possibility of finding such unique S–R connections in *fact* is a prophylactic against using m and f as interchangeable terms.

From the analysis of meaning presented previously and from the Zipf-Thorndike findings, one may infer that verbal stimuli acquire m-value as some increasing function of their frequencies of occurrence (n). Symbolically, this relationship may be written

$$m = \Phi(n). \qquad [2]$$

If f were next shown to be some increasing function of n; i.e., if

$$f = \Psi(n), \qquad [3]$$

then it validly may be concluded that f is therefore some increasing function of m:

$$f = \Gamma(m). \qquad [4]$$

One may now proceed to test the deduction described in equation [4] by determining the empirical relationship between the meaning index (m) and judged frequency or familiarity (f). In Fig. 4 are plotted the results of a comparison of the f-scale with the m-scale. Familiarity (f) is shown as an increasing curvilinear function of the meaning index (m). The equation describing this relationship, determined by a least squares solution, is

$$f = 6.04(1 - 10^{-0.007m}) + 0.28. \qquad [5]$$

The index of correlation ρ, computed from the reduction line, has a value of 0.92.

It was noted in (10) that the response-frequency distributions for the stimuli on the m-scale were positively skewed for low values. In view of this fact, medians were computed in Fig. 4.

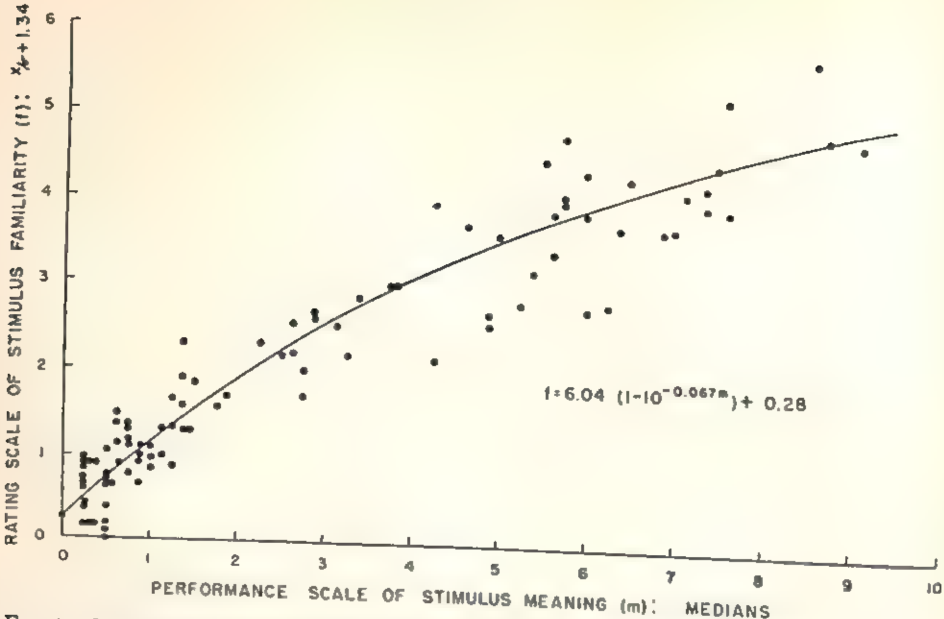


FIG. 4. Scattergram of the relationship between familiarity (f) and meaning (m). Ordinate values denote graded dichotomies scale values from ratings on Familiarity Schedule ($N = 200$). Abscissa values denote corresponding median frequencies of continued association to each of 96 dissyllables within 60 sec. ($N = 119$). The equation and fitted curve were derived from the method of least squares.

This measure of central tendency affords a more valid estimate of the zero point of the m-scale for the present purposes. It is clear that f is a negatively accelerated positive function of m. The Γ-function is neither one of identity nor of proportionality, a fact consistent with the preceding logical analysis.

The empirical law represented in equation [5] does not imply that m *determines* f in any sense of strict implication, such as that denoted by the assertion: "If and only if $m >$ zero, then $f >$ zero." It is only to say that f conveniently may be represented as a correlate of m. Obviously, there might be a third variable with which m and f are co-correlated; e.g., n, as suggested above. Hence, the hypothesis may be entertained that verbal stimuli acquire the attributes m and f as joint functions of frequency of stimulation.

The stimulus frequency hypotheses represented by equations [2] and [3] above are not regarded as verified, but as only tenable pending further tests. One attempt to falsify them might consist in experimentally establishing various m-values for a set of *new* verbal stimuli in a group of *S*s, then having those *S*s later rate the same stimuli for familiarity. This would provide a check on an hypothesis the evidence for which is at present only indirect; i.e., this analysis is concerned with *individual-difference* as distinguished from *experimental* variables. The suggested test would involve the "creation" of new stimulus attributes rather than assaying those already "built into" the *S*s as a consequence of their prelaboratory experiences. The test might also be extended to nonverbal *S–R* situations, such as classical conditioned responses or perceptual-motor learning tasks.

In this context, the procedure of endowing stimuli with the properties of meaningfulness (m) or familiarity (f)

may constitute one unambiguous definition of Thorndikian "identifiability," which in turn may be related to such current notions as "pre-differentiated structure," "distinctiveness," "cue-value," and "recognizability." Noble (11) has shown, for example, that the variable m plays an important role in verbal learning.

It should be noted that the rationale governing the graphical representation of the data in Fig. 3 in the form $f = \Gamma(m)$ is that of the distinction Bergmann and Spence (2) have made between *strict behavioristic* terms and those which are merely *behavioristic*. This is further related to their distinction between *S–R* and *R–R* laws (cf. also 13). Adopting their symbolism, one writes the meaning-familiarity function $R_1 = \Gamma(R_2)$. The empirical relationship is expressed in this form for purely methodological reasons. It amounts to considering the regression of test scores on criterion scores. In other words, a distinction is made between ratings and performance. It should be emphasized that the m-scale provides fundamental measures (average frequencies) with an absolute zero point, while the f-scale yields derived measures (x/σ transformations) having an arbitrary origin. For any accurate quantitative purposes, therefore, the m-scale possesses certain advantages. It should also be recalled that an m-value is obtained by one simple, independent operation (mean or median) upon a single distribution of responses, while an f-value involves many complex, dependent operations upon (preferably) all the distributions of a set.

SUMMARY AND CONCLUSIONS

This paper has presented a theoretical-experimental analysis of the attribute of familiarity in verbal stimulus material. A word list of 96 dissyllables consisting of nouns and paralogs was presented to a sample of 200 U.S.A.F. recruits in order to establish a quantitative scale for this attribute. The results of this analysis were as follows:

1. Familiarity was formally defined as a stimulus attribute which is some increasing function of the frequency of occurrence of a given stimulus.

2. An index of familiarity (f) was operationally defined in terms of a five-point graphic rating schedule.

3. An equal-interval rating scale of f-values was developed which exhibited a range extending from 0.00 to 5.66. A test of internal consistency indicated an average error of 0.019 in reproducing the original data from the scale values.

4. Certain characteristics of the f-scale relevant to the theory of measurement were discussed.

5. The hypothesis was proposed that stimuli also acquire the attribute of meaning (m) as some monotonic function of frequency, and that, therefore, f should be positively related to m.

6. An attempt was made to determine the exact functional relationship between m and f. The equation was

$$f = 6.04(1 - 10^{-0.087m}) + 0.28.$$

7. The significance of these findings for research in verbal and in perceptual-motor learning was discussed.

REFERENCES

1. ATTNEAVE, F. A method of graded dichotomies for the scaling of judgments. *Psychol. Rev.*, 1949, **56**, 334–340.
2. BERGMANN, G., & SPENCE, K. W. The logic of psychophysical measurement. *Psychol. Rev.*, 1944, **51**, 1–24.
3. EDWARDS, A. L. Psychological scaling by means of successive intervals. *Bull. Psychom. Lab. Univer. Chicago*, 1951, No. 69.
4. GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
5. HAAGEN, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27**, 453–463.

6. HEVNER, KATE. An empirical study of three psychophysical methods. *J. gen. Psychol.*, 1930, **4**, 191–212.

7. MCGEOCH, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.

8. MOSIER, C. I. A modification of the method of successive intervals. *Psychometrika*, 1940, **5**, 101–107.

9. NOBLE, C. E. Absence of reminiscence in the serial rote learning of adjectives. *J. exp. Psychol.*, 1950, **40**, 622–631.

10. NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430.

11. NOBLE, C. E. The rôle of stimulus meaning (m) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437–446.

12. SAFFIR, M. A. A comparative study of scales constructed by three psychophysical methods. *Psychometrika*, 1937, **2**, 179–198.

13. SPENCE, K. W. The nature of theory construction in contemporary psychology. *Psychol. Rev.*, 1944, **51**, 47–68.

14. THORNDIKE, E. L. On the frequency of semantic changes in modern English. *J. gen. Psychol.*, 1948, **39**, 23–27.

15. THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Columbia Univer. Press, 1944.

16. THURSTONE, L. L. Psychophysical analysis. *Amer. J. Psychol.*, 1927, **38**, 368–389.

17. UNDERWOOD, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.

18. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

19. ZIPF, G. K. *Human behavior and the principle of least effort.* Cambridge, Mass.: Addison-Wesley, 1949.

[MS. received February 11, 1952]

Psychological Review
Vol. 60, No. 2, 1953

HOW ARE MOTIVES LEARNED?

JOHN P. SEWARD

University of California, Los Angeles

One of psychology's basic problems is to account for the variety, no less than the uniformity, of human wants. At one time the instinctivists seemed to hold the key; at another, the advocates of conditioning. When both of these doctrines in their early forms proved inadequate, students of behavior still faced the problem of bridging the gap between the bodily needs of the infant and the adult's desires to succeed or fail, to help or hurt, to be loved or hated.

Equally dissatisfied with the instincts of the hormic and psychoanalytic schools and with the conditioning of biological drives, Allport proposed the "functional autonomy of motives" as "a declaration of independence for the psychology of personality" (1, p. 207). But since he neglected to write a constitution for the newly liberated discipline, his principle did little to solve the problem. His most explicit statement was a paraphrase of Woodworth's original suggestion that in the course of activity *mechanisms* become *drives*; "activities and objects that earlier in the game were *means* to an end now become *ends* in themselves" (1, p. 195). How this transformation occurs was left untouched.

Another writer to grapple with the problem was Murphy (25). Explicitly rejecting functional autonomy, he broadened the base of unlearned drives and introduced the concept of *canalization* to represent the process by which they become focused on specific satisfiers. But to substitute a canalized drive for an autonomous motive is hardly explanatory. Both concepts belong in the category of "nominal theories" (32).

A third formulation that falls in the same class is Tolman's (33) drive-conversion diagram. When biological drives are frustrated, they are converted into social drives or techniques (e.g., self-assertive or collective) by the mechanism of learning and the dynamism of fixation. A technique is fixated to the extent that it persists in situations to which it is inappropriate. But to this extent it seems to become a drive, or at least equivalent to Allport's autonomous motive. If we only knew the conditions of fixation! With characteristic candor Tolman called his diagram "a scheme more for the asking of important questions than for the presenting of final answers" (35, p. 227).

A search for more precise conditions of motive learning calls for a distinction between learned drives and learned rewards. If a *drive* is learned it means that an originally neutral stimulus will instigate activity even when the primary drive conditions are removed. But if a *reward* is learned it means simply that activity otherwise instigated is *directed* toward an originally neutral stimulus. In the first case motivation is at least temporarily free of its primary source; in the second case it may or may not be. Whether reward learning implies some degree of drive learning as well is, in fact, the chief problem of this paper.

A second distinction empirically related to the first is between *appetites* and *aversions*, to borrow Tolman's terms. For our purpose an aversion may be defined as a motive started by the presence of a certain stimulus and ended by its removal. An appetite is a motive *ended* by commerce with a certain stimulus, but what starts it we cannot say without knowing more about the part played by learning. On the whole,

as we shall see, aversions lend themselves more directly to drive learning, while appetites seem more prone to produce and be produced by the learning of rewards. This difference is, I believe, too often glossed over. Tolman, for example, in his latest discussion of the problem, described *positive* and *negative cathexes* as "attachments of specific types of final positive goal-object, or of final negative 'disturbance-object' to basic drives" (34, p. 144). (The basic drives are presumably appetites and aversions, respectively.) Although Tolman treated them as strictly analogous, on closer inspection the parallel breaks down. For one thing, response to the positively cathected goal-object comes *after* the drive is in force, while the negatively cathected disturbance-object comes *before* the drive and serves to arouse it. For another, positive cathexes are reinforced, for Tolman, by need-reduction, but negative ones are reinforced by "pain or some other type of noxious physiological state" (34, p. 147). Thus it is not at all clear that "cathexis learning" is a unitary construct either in method or result.

Mowrer (22) recognized this potential split without pursuing its implications. In asserting that drives and other visceral responses are learned by contiguity and skeletal ones by the law of effect, he emerged, as he put it, with one principle for the making of problems and another for solving them. Since his dichotomy grew out of experiments involving fear and its reduction, aversions fit readily into his scheme. But appetites are left in an ambiguous position. Is goal anticipation problem making or problem solving? Drive increasing or drive reducing? Conditioned or "reinforced"? In discussing the Pavlovian experiment Mowrer suggested that the salivary CR be considered "an indication, perhaps a veritable part of the mechanism, of the secondary motivation known as 'appetite'" (24, p. 257). But even as such it may mean merely that the dog is hungry and expects food, not that he has acquired a desire for it whether hungry or not.

Hull expresses similar uncertainty with respect to the systematic status of incentives. On the one hand he developed the functions of the fractional antedating goal reaction and its proprioceptive consequent in purposive behavior (13, 14, 15). On the other hand he made room for secondary motivation through the reinforcement of connections between neutral stimuli and response-produced drive stimuli. In his latest publication he writes: ". . . the understanding of . . . secondary motivation is rather obscured by the fact that its relationship to incentive motivation . . . and to the associated mechanisms of anticipatory goal reactions . . . has not yet been clarified" (16, p. 24).

A major attempt to apply a drive-learning formula to appetites was made by Anderson (2). He proposed that the neural drive mechanism, at first aroused by internal conditions, becomes externalized, i.e., activated by external situations in which it has been repeatedly satisfied. Anderson's theory is important both because of its experimental fertility (3, 4, 5) and because it may even be true. As it stands, however, it illustrates some of the difficulties we must face. What has the *satisfaction* of a drive to do with its externalization? How is a drive, externalized to a goal object, aroused in its absence? How does an externalized drive cause persistent behavior in a satiated organism but not consumption of the goal object?

Miller, who leads the experimental attack on the learning of motives, has recently summarized research in this area (21). His review bears out our impression of an asymmetrical development. Writing on "learnable drives and rewards," Miller devoted 18 pages to the

acquiring and eliminating of fears. Of eight pages on food hunger, six dealt with the learning of rewards, including a table of 21 studies, the other two with the learning of drive. Of the experiments cited in these two pages, only three or four could be considered at all definitive. If fear and hunger are typical we must conclude that aversions yield plenty of evidence for drive learning and that appetites show plenty of reward learning, but that evidence for the learning of appetitive drives is scanty.

One reason for this discrepancy may lie in a difficulty of experimental technique. To demonstrate secondary motivation, the influence of the primary drive must be ruled out. In aversions this is easy; one need only turn off the shock. But how can the effect of hunger be eliminated? The most obvious way is to feed the animal to the point of refusal. To reach this point, however, it may not be enough to reduce the need for food to zero, since we must counteract not only a metabolic deficit but also any acquired drive to eat. But this means removing a part of the very process we seek, the remainder of which we must try to measure by some other response. This argument is not entirely circular. Most experiments on the strength of food-getting responses as a function of hunger (12, 17, 27) agree that the curve shows a break at about two hours of deprivation, rising steeply below that point and only gradually above it. These data suggest that satiation introduces a factor (full stomach?) that inhibits activity associated with food. Until other means of control are devised, learned appetitive drives may well continue to elude us.

A second reason why aversions may be easier to condition than homeostatic drives was suggested by Miller (21, p. 463) and has to do with their relative abruptness of onset. To be conditioned, a response must be evoked in the presence of the CS. Hull states explicitly that for secondary motivation to develop, neutral stimuli must be "associated with the *evocation* of a primary or secondary drive" (italics mine) (16, p. 25). Escape drives can be aroused by noxious agents applied suddenly in conjunction with neutral stimuli; hunger or thirst, on the other hand, develops slowly. The same reason may explain why appetites show readier conditioning of consummatory than of drive-producing responses.

These and possibly other hypotheses must be explored before the data can be finally evaluated. Meanwhile it may help to face squarely the issue before us. Experiments strongly imply that the chief sources of motivation are a few bodily needs, a variety of ways of satisfying them, and a large number of fears. Common experience, on the other hand, suggests that the bulk of our activity is directed toward things we want rather than away from things we don't want, and that many of the things we want have little to do with bodily needs. How can these conflicting lines of evidence be reconciled?

Several ways are open. An extreme behaviorist, for example, will limit his materials to experimental observations and closely related constructs and will stretch them as far as he can. The theory he fashions may run something like this: The only unlearned drives are bodily discomforts and the only learned drives are fears. The apparent motive force of positive desires is illusory; a cigarette craving is anxiety, sociability is a fear of solitude. Goals are not really motives at all, but habits acquired by drive reduction, activated by drives, and capable of providing secondary reinforcement for other habits. The few cases in which subgoals have been found to reinforce habits in satiated animals may be explained by Mowrer's concept of *drive-fear* (23). A well-fed, warmly

clad wage earner seeking a raise is driven by hunger-fear or cold-fear. At the other extreme a "personalistic" psychologist may describe the growth of human aspirations with little regard for laboratory results, on the assumption that these are largely irrelevant to personality dynamics.

The theory in the next section is presented as an attempt to account for the experimental evidence on learning drives and rewards and also, it is hoped, as a step toward the identification of other factors for which experimental evidence is not yet available.

Theory of Motive-Learning

In essence I suggest that acquired drives fall into two classes, secondary and tertiary, depending on how they are acquired. Secondary drives result from the conditioning of primary drive components to external stimuli. Tertiary drives result from the blocking of learned responses, especially responses by which primary or secondary drives have been reduced. Which kind of drive is acquired depends on whether circumstances favor the conditioning of drives or the learning of goal responses. There is no *necessary* connection between this classification and that of primary drives into aversions and appetites. As Miller has pointed out, however, the onset of aversions is typically sudden; that of appetites, gradual. It seems reasonable to suppose that in order to be learned a response must involve a somewhat abrupt change of excitation. We should, therefore, expect aversions to yield secondary drives supplemented by tertiary, while appetites, *unless their onset is sudden*, should lead almost exclusively to tertiary drives.

The most convincing evidence of acquired motivation is the learning of a new response in the absence of primary drive. Our theory should therefore enable us to deduce this finding both where conditions favor drive-learning, as in aversions, and where they favor reward learning, as in appetites. To be of value the theory should also "explain" other pieces of evidence, generate additional hypotheses to test, and permit expansion to include a wider range of behavioral phenomena. With these criteria in mind I submit the following definitions and postulates.

Definitions

D1. *Surrogate response* (*rs*). A central neural process by means of which a stimulus (S) evokes a response (R).[1] Strength of R, as measured by amplitude, latency, relative frequency, or extinction rate, is a function of intensity of *rs*. Under special conditions (to be stated in other definitions and postulates) *rs* can function independently of its primary S or R.

D2. *Conditioning*. A rs_1 is said to be conditioned to rs_2 when rs_1 acquires a *functional connection* with rs_2. Such a connection is called *adequate* in so far as rs_1 is actually followed by rs_2; it is called *potential* in so far as arousal of rs_2 must await the aid of other S's or *rs*'s.

D3. *Primary drive stimulus* (S_D). A S producing or produced by a homeostatic imbalance.

D4. *Drive* (rs_D). A *rs* first aroused by S_D or by postulate 4 and able to strengthen the *rs*'s to a hierarchy of R's as an increasing function of its own intensity.

D5. *Primary goal stimulus* (S_G). A S producing or produced by a homeostatic balance (G) with respect to a particular need.

D6. *Goal surrogate* (rs_G). A *rs* first aroused by S_G or by postulate 4 and varying in intensity as a negative linear function of the intensity of rs_D.

[1] The term *stimulus* and *response* are used for convenience to refer to patterns of receptor and effector activity of varying complexity.

D7. *Subgoal response* (R_g). A R closely and consistently followed by an abrupt increase in intensity of rs_G.

Postulates and Corollaries

P1. *Association.* If rs_1 is active when rs_2 is aroused, rs_1 becomes conditioned to rs_2. Strength of conditioning is a positive function of the number of concurrences and a negative function of the time interval between $S_1 - R_1$ and $S_2 - R_2$.

C1. *Secondary drive* (rs_{D_2}). A rs accompanied by an abrupt increase in intensity of rs_D becomes conditioned to it and its S becomes a *secondary drive stimulus* (Srs_D).

C2. *Secondary reward.* A rs accompanied by an abrupt increase in intensity of rs_G becomes conditioned to it and its S becomes a *secondary goal stimulus* (Srs_G).

P2. *Activation.* The strength with which rs_2 is activated by rs_1, as a result of a potential or adequate connection, is a positive function of the (a) strength of connection, (b) present intensity of rs_1, (c) intensity of rs_2 during conditioning.

P3. *Facilitation.* An increase in intensity of rs_G strengthens the rs of any R in progress as a positive function of the amount and rate of increase.

P4. *Tertiary drive* (rs_{D_3}). If rs_1 has a strong potential connection with rs_2, arousal of rs_1 produces a drive (rs_{D_3}) varying in intensity as a positive function of strength of activation of rs_2. Arousal of rs_2 reduces the drive, the amount of reduction varying positively with the intensity of arousal.

P5. *Summation.* If two rs's conditioned to the same rs_3 are simultaneously active, they arouse rs_3 with a total intensity which is greater than that evoked by either one alone.

A word must be said about the two pivotal concepts, *drive* and *goal surrogate*. Although given the status of rs and thus subsumed under P1, they have certain special properties. The rs_D is a central excitatory state, while rs_G is a state of diminished excitation negatively related to it. Although it may ultimately prove unnecessary to retain both constructs, at present it seems to me to simplify the theory. For one thing, it avoids the awkward notion of conditioning the reduction or termination of a process. A rather specific relationship is envisaged here: i.e., to the extent that two rs_D's are discriminable the corresponding rs_G's are likewise distinct; to that extent rs_{D_1} and rs_{G_2} vary independently of each other. The rs_G produced by eating, for example, reduces only slightly if at all the rs_D of sex and is but little reduced thereby. Hunger and thirst, on the other hand, are so closely related that it would not be safe to make the same assumption regarding them (8, 36). Some drives, too, may have proprioceptive components in common and, at high intensities, visceral ones as well. Such overlapping complicates the problem of isolating learned drives. In so far as it can be made, the assumption of independence is important for us because of the following consequence: *if a given $rs_D = 0$, the strength of its rs_G cannot be further increased.*

Despite certain changes the postulates have a somewhat parasitic relation to Hull's system and do not pretend to be self-sufficient. My chief departures from Hull are two: (a) Drive reduction is believed to strengthen, not the conditioning of R's, but the R's themselves; (b) the concept of *surrogate response* replaces his response-produced stimulus and is given a focal position. P1, for example, may be recognized as a modified version of Hull's postulates 4 (habit formation) and 8 (delay in reinforcement) (16), restated in terms of rs's and without reference to reinforcement. P2 is a partial restatement of his postulate 9 (constitution of reaction potential).

"Strength of conditioning" and "strength of connection" are synonymous and are equivalent to Hull's *habit strength*, while "intensity of *rs*" is the counterpart of his *reaction potential*. The function of drive reduction in the present theory is embodied in P3. Note that it affects activity in progress, not strength of connections, another way of saying that "reinforcement" is here thought of as a principle of performance rather than of learning.[2]

P4 attempts to state the basic mechanism whereby a goal acquires motivating power. A similar idea has been formulated by Whiting in an unpublished manuscript cited by Miller (21); P4 is, therefore, not original except in its present form. Emphasis should be placed on the concept of *potential* as distinguished from *adequate* connection; i.e., a habit becomes excitatory when it is started but not finished.[3] As it stands, the postulate is exceedingly primitive. It needs first of all a statement of the conditions that result in potential connections. One example is a R that cannot be completed without an external S; one cannot eat without food or press a bar that isn't there. Another is a time interval between S_1 and S_2; presumably the shorter the time the more adequate the connection, a relation that may be reduced, as Spence (31) has suggested, to distance on a generalization gradient. A second problem concerns the quantitative relation between activation of rs_2 and strength of rs_{D_3}. We are familiar with the heightening of tension as we approach a goal, yet we should expect rs_{D_3} to be reduced by the more adequate arousal of rs_2. A solution might be to break down purposive behavior into a series of subgoals; as each one is attained, it reduces its own rs_{D_3} but adds potential leading to a later stage. Such possibilities must be left for future development.

Deductions

1. *Response selection under learned drive.* So far the above propositions, with the help of a few others, have led to the more or less formal deduction of nine theorems dealing with one or another aspect of acquired drives. Only those directly concerned with R selection under rs_{D_2} and rs_{D_3} will be reproduced here; others will be merely indicated along with the supporting evidence.

First we need a general derivation of the phenomena usually attributed to selective learning. Theorem 1 is an attempt to meet this need.

T1. *Response selection. Given:* On one or more occasions an organism (O) under drive (rs_D) in the presence of S makes a certain subgoal response (R_g).

To prove: On recurrence of rs_D and S the probability of R_g will be greater.

Proof: 1. Let R_x be any other R *not* R_g made in the presence of rs_D and S. (For simplicity let R_x and R_g be equivalent in other respects; e.g., availability, effort, previous conditioning.)

2. R_g, but not R_x, is followed by rs_G (D7).

3. rs_g is conditioned to rs_G more strongly than is rs_x (P1, C2).

4. S, by way of rs, is conditioned to both rs_x and rs_g, as is also rs_D (P1).

5. The next time S and rs_D arouse rs_g they will induce a greater intensity of rs_G than when they arouse rs_x (Step 3, P2).

6. In turn, rs_G will strengthen rs_g more than rs_x (P3).

7. The frequency of R_g relative to

[2] In view of the inverse relation between rs_D and rs_G it is tempting to suggest a counterpart of P3; viz., an increase in intensity of rs_D weakens the rs of any R in progress. Since this hypothesis is not needed here, its possibilities will not be developed.

[3] Compare with Allport's suggestion that it is a habit-in-the-making that becomes autonomous (1, p. 204 f.).

R_x—i.e., its probability—will be increased (D1). Q.E.D.

To go from T1 to the special case of R selection under rs_{D_2} is only a step, as shown in theorem 2:

T2. *R selection under rs_{D_2}. Given:* (1) A neutral S is accompanied one or more times by an abrupt increase of rs_D. (2) S is later presented alone and removed when O makes a certain R.

To prove: On recurrence of S the probability of R will be greater.

Proof: 1. After one or more conjunctions with a sudden increment of rs_D, S alone arouses rs_D (C1).

2. Removal of S results in an abrupt increase of rs_G (D6).

3. R thus becomes R_g (D7).

4. The conditions of T1 are thus fulfilled and R will occur with increased probability in the presence of S.

Q.E.D.

Brown and Jacobs (7) provided the experimental model for this theorem. After repeatedly giving rats a neutral S paired with shock, they presented S alone and removed it when the animal jumped a barrier. They found a progressive decrease in latency of jumping, a result that could just as easily be deduced as an increase in probability (the one, in fact, implies the other).

Closely resembling the Brown-Jacobs experiment was its forerunner, Miller's (20) classical demonstration of fear learning. In that study, however, on training trials a running R regularly led to escape from shock. When this R_g was prevented by closing the door, it could have generated rs_{D_3} as the motive for learning a door-opening R. Indeed, it was to eliminate this alternative that Brown and Jacobs made their changes. Since Miller's experiment, for all its value, could be explained by either rs_{D_2} or rs_{D_3}, it was crucial to neither.

Our next objective, therefore, is to derive a more rigorous test of rs_{D_3}. T1 provides for the selection of R_g under drive. But to free rs_{D_3} from other drive effects we must be able to predict that when S is later presented without either primary drive (rs_{D_1}) or rs_{D_2}, it will still tend strongly to arouse R_g. Note that if we assume a minimum[9] of drive interaction, and $rs_D = 0$, rs_G cannot be increased and therefore cannot facilitate rs_g. With the help of P2, however, the deduction is readily made, as shown by theorem 3:

T3. *Response persistence. Given:* the same conditions as in T1, save that R_g must occur on *more than one occasion* during training.[4]

To prove: On recurrence of S in the absence of rs_D, the probability of R_g relative to other R's will be greater.

Proof: 1. Again let R_x be any other equivalent R not R_g.

2. After one trial under rs_D, due to facilitation by rs_G S is subsequently conditioned to a greater intensity of rs_g than of rs_x (Steps 2 to 5, T1).

3. On later recurrence without rs_D, S will still arouse rs_g more intensely than rs_x. (By the last provision of P2 the intensity of rs aroused is a positive function of the intensity of rs conditioned.)

4. S will evoke R_g more readily than R_x (D1). Q.E.D.

T3 is not without evidence. Teel and Webb (32) trained rats to food on one side of a T after varying amounts of deprivation. Every day after two deprived trials to each side they gave a pair of satiated trials. The percentages of correct choices under deprivation and satiation were practically the same throughout training. Seward and Handlon (28) gave thirsty rats an equal number of forced trials to water on one side

[4] The reason for this proviso will become clear in step 2 of the proof. Incidentally, since a reinforcement learning theory does not make this requirement, we have here a possible technique for putting the two hypotheses to a crucial test.

of a choice point and no water on the other. When satiated for water, the rats still showed a significant shift of preference to the water side.

Consider another implication of the theorem. Let us assume that rs_G can facilitate only under rs_D but that its effect persists when rs_D is removed. Let us further assume that two R's to S are both followed by Srs_G, but one (R_g) is made under rs_D and the other (R_x) is not. It follows that when S is presented without rs_D, R_g will be preferred to R_x. Seward, Levy, and Handlon (30) tested this implication. They forced rats equally often to water on both sides of a choice point, but thirsty to one side and satiated to the other. Tested under satiation the rats shifted their preference to the thirsty side.

We are now ready for an attempt to use R selection as a criterion of rs_{D_3}.

T4. *R selection under rs_{D_3}. Given:* (1) An O under rs_D in the presence of S_1 makes more than once a certain R_g that cannot be completed without S_g. (2) S_1 is later presented one or more times without rs_D, and with S_g presented only when O makes a certain R_c.

To prove: On recurrence of S_1 the probability of R_c will be greater.

Proof: 1. O comes to the test trials with a relatively strong tendency for rs_1 to activate rs_g (T3).

2. Withholding S_g creates the conditions for rs_{D_3} (D2, P4).

3. O makes a number of R's (D4). Let R_c be one of these and let R_x be any other R not followed by S_g.

4. The following connections are formed: $rs_1 - rs_x$, $rs_1 - rs_c$, and $rs_c - rs_g$ (D1, P1).

5. On subsequent trials, when rs_1 arouses rs_c, rs_g is evoked with greater intensity than when rs_1 arouses rs_x (P5).

6. But this added intensity reduces rs_{D_3}, thus increasing rs_G (P4, D6).

7. The conditions of R selection are thus provided, resulting in a greater probability of R_c than R_x (T1).

Q.E.D.

As already said, evidence for rs_{D_3}, free of rs_{D_2}, is most likely to be found where rs_{D_1} is a slow-rhythm appetite. But the experimental demonstration of acquired drive based on appetites, possibly due to the satiation effect already mentioned, has proved difficult. In a study so far only briefly reported Miller (19) trained two groups of rats to run from a white to a black box, one to escape shock and the other to allay hunger. He then compared their performances without rs_{D_1}. Not only did the shocked rats maintain their R_g longer than the fed ones but they learned a new R, whereas the other group did not. Under the usual experimental conditions rs_{D_3} seems definitely weaker than rs_{D_2}. Miller's report contained one bit of evidence for rs_{D_3}. He trained several groups of rats to run through a door to food, then tested the R_g under satiation. One group that had been blocked by the door for varying delays during training was the slowest to extinguish. The present theory would predict this result on the ground that blocking the door would generate rs_{D_3}, which would be perpetuated on test trials through the conditioning of a stronger rs_g.

Evidence from maze studies is more favorable. Some years ago Anderson (4) found indications that rats trained hungry and rewarded in one maze did better in a new maze, even though satiated, than a control group previously trained under less favorable drive-incentive conditions. He interpreted his results in terms of drive externalization. If his promising data were confirmed, I should prefer to say that training conditioned strong potential rs_gs to maze S's and that these rs_g's provided rs_{D_3} for the selection of R's leading to their completion. To predict Anderson's findings, T4 would have to be adapted only to

the extent of providing for stimulus generalization from one maze to another.

More recently MacCorquodale and Meehl (18) put food and water in the same endbox of a T. Rats satiated for both objects nevertheless chose that side significantly more than the empty one. In an unpublished study Myers (26) confirmed this result impressively. Putting a rat as a social incentive on each side of a T, but food on only one, he gave satiated rats two trials a day, one free and one forced opposite. In 12 days they increased their food-side choices progressively and significantly. When the food was then shifted to the other side for 26 days, they gradually reversed their preference, 21 out of 26 rats reaching a criterion of 8/10 "correct."

These two studies do not precisely fit the paradigm of T4 in that S_1, the choice point, was not combined with S_g, the incentive, during an initial training period under rs_{D_1}. "Training" presumably took place in O's feeding history, during which the rs_g's involved in seeing and smelling food became strongly conditioned to—and thus facilitated by—the rs_G produced by eating. The results suggest that if S_1 is conditioned to a rs_g that is already strong enough, it will produce rs_{D_3} without further help from primary drive or reward.

Granting that food and water can bring about new learning in the absence of the original drives, we may ask if rs_{D_3} learning is confined to such "canalizers" or if it can be mediated by other objects not so intimately related to drive reduction? Anderson's study suggested the latter possibility. Seward and Levy (29) tested the hypothesis that rats rewarded in one compartment but not in another would show a preference, even when satiated, for the compartment that had previously contained the reward. They first trained thirsty rats in a runway to two different boxes, one containing water and the other none. (In contrast with the two studies just discussed, it was here found necessary to associate runway Ss with endbox Ss under drive-reward conditions in order to get positive results.) Later the rats were satiated and run in a T with one box on either side. Test trials gave clear and consistent evidence of the effect of training on choices, with highly significant differences in favor of the positive box. Moreover the presence or absence of food and water in the boxes during tests was apparently without effect. Its importance during *training*, however, was further brought out by interpolated runway trials under thirst. When water was removed from *both* boxes on these trials, the difference between them in satiated test choices dwindled rapidly. Apparently a need-reducing R_g that has become the basis of rs_{D_3} may lose that function if, when the original need is present, R_g no longer satisfies it.

2. *Other evidence.* Various signs of acquired drive in the experimental literature lend themselves to interpretation in terms of the present theory, but here they can be only briefly mentioned. (*a*) *Goal-instigated activity* was strikingly demonstrated by Geier and Tolman (10, 11). They devised a method of measuring *frustration-tension*, a construct that seems identical with tertiary drive. (*b*) *Response persistence* after rs_{D_1} is removed does not necessarily imply an acquired drive since it may be due to $_SH_R$ combined with an irrelevant D. But Farber (9) showed that R was abandoned more quickly when rs_{D_2} had been extinguished than when it had not. And Brogden (6) provided parallel evidence for rs_{D_3} when he retarded extinction of a flexion R in satiated dogs by "rewarding" it with a pellet.

UNSOLVED PROBLEMS

In the last section a theory of acquired drives was presented and ex-

amined in the light of experimental evidence. As anticipated, all the evidence for rs_{D_2} came from experiments using shock, while support for the concept of rs_{D_3} was drawn exclusively from experiments using hunger or thirst. This dichotomy arose as a natural result of the way the two concepts originated: the theory of secondary drive was based primarily on situations involving fear; that of tertiary drive grew out of studies on appetites. The question remains whether this is a distorted view or whether it represents a true distinction between appetites and aversions. How much of what we have called rs_{D_3} is actually "externalized" hunger? How much of fear is a desire for safety?

A way to find out is suggested by the experiments of Miller (20) and of Brown and Jacobs (7) already cited. The first of these studies provided conditions for acquiring both rs_{D_2} and rs_{D_3}; the second, for rs_{D_2} alone. If two such experiments were performed under strictly equated conditions the relative "weights" of the two sources of drive might be determined. And if a similar pair of experiments was carried out substituting hunger for shock, a direct comparison could be made in this respect between a typical aversion and a typical appetite.

Supposing that appetites and aversions do differ as suggested, how are we to explain it? Is Miller correct in supposing that relative abruptness of onset is the distinguishing feature? If so, a suddenly induced hunger should give rise to rs_{D_2}, while a gradually developing aversion should not; on the other hand, such an aversion, if suddenly relieved, should yield rs_{D_3}. All we need is a way of reversing the usual time relations.

Even if we accept the idea that an appetite generates rs_{D_3}'s we must still face the question: why are they so much more elusive than fears? There are several possibilities to be explored. One factor already mentioned is the counteracting of acquired motives by acute satiation of the primary need. A method is required for isolating goal-instigated behavior without stuffing the animal. A second hypothesis is that goal and subgoal responses must somehow be made stronger before their incipient arousal can act as a drive. Perhaps this could be done by increasing the amount of training, or strength of rs_{D_1}, or the number of different needs satisfied, or the effort required to make R_g. Thirdly, it is possible that rs_{D_3} is a fairly high-level cerebral achievement; as such it would probably be easier to develop in animals with bigger brains than rats.

SUMMARY

This paper starts with a dilemma. On the one hand human activities seem to be largely instigated by goals that have little to do with bodily needs. It has been suggested that primary drives are conditioned to external objects, which come to arouse them independently of tissue disturbance. On the other hand experiments with animals have shown that, although fear is easily learned, the appetitive drives are not. At the same time goal-associated responses are readily "cathected" in connection with these drives. How can the facts of life and the facts of the laboratory be reconciled?

A theory is proposed, the gist of which is that there are two ways by which motives can be learned. Secondary drives are conditioned directly to neutral stimuli; tertiary, or goal-instigated, drives are touched off by the obstruction of a learned response. Owing, perhaps, to a difference in characteristic rates of onset, aversions give both kinds of drives but appetites give mainly the tertiary type. To make possible a more precise integration of theory and experiment a set

of definitions and postulates is offered and several illustrative theorems are derived. These are discussed in relation to available evidence.

But the initial dilemma remains to challenge the ingenuity of future experimenters.

REFERENCES

1. Allport, G. W. *Personality: a psychological interpretation.* New York: Holt, 1937.
2. Anderson, E. E. The externalization of drive. I. Theoretical considerations. *Psychol. Rev.*, 1941, **48**, 204–224.
3. Anderson, E. E. The externalization of drive: II. The effect of satiation and removal of reward at different stages in the learning process of the rat. *J. genet. Psychol.*, 1941, **59**, 359–376.
4. Anderson, E. E. The externalization of drive: III. Maze learning by non-rewarded and by satiated rats. *J. genet. Psychol.*, 1941, **59**, 397–426.
5. Anderson, E. E. The externalization of drive. IV. The effect of pre-feeding on the maze performance of hungry non-rewarded rats. *J. comp. Psychol.*, 1941, **31**, 349–352.
6. Brogden, W. J. Non-alimentary components in the food-reinforcement of conditioned forelimb-flexion in food-satiated dogs. *J. exp. Psychol.*, 1942, **30**, 326–335.
7. Brown, J. S., & Jacobs, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747–759.
8. Estes, W. K. Generalization of secondary reinforcement from the primary drive. *J. comp. physiol. Psychol.*, 1949, **42**, 286–295.
9. Farber, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111–131.
10. Geier, F. M. The measurement of tension in the rat. *J. comp. Psychol.*, 1942, **34**, 43–49.
11. Geier, F. M., & Tolman, E. C. Goal distance and restless activity. I. The goal gradient of restless activity. *J. comp. Psychol.*, 1943, **35**, 197–204.
12. Horenstein, B. R. Performance of conditioned responses as a function of strength of hunger drive. *J. comp. physiol. Psychol.*, 1951, **44**, 210–224.
13. Hull, C. L. Goal attraction and directing ideas conceived as habit phenomena. *Psychol. Rev.*, 1931, 38, 487–506.
14. Hull, C. L. Mind, mechanism, and adaptive behavior. *Psychol. Rev.*, 1937, **44**, 1–32.
15. Hull, C. L. The concept of the habit-family hierachy and maze learning. *Psychol. Rev.*, 1943, **41**, 33–52, 134–152.
16. Hull, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.
17. Kimble, G. A. Behavior strength as a function of the intensity of the hunger drive. *J. exp. Psychol.*, 1951, **41**, 341–348.
18. MacCorquodale, K., & Meehl, P. E. "Cognitive" learning in the absence of competition of incentive. *J. comp. physiol. Psychol.*, 1949, **42**, 383–390.
19. Miller, N. E. Experiments on the strength of acquired drive based on hunger. *Amer. Psychologist*, 1947, **2**, 303. (Abstract.)
20. Miller, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.
21. Miller, N. E. Learnable drives and rewards. In Stevens, S. S. (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 435–472.
22. Mowrer, O. H. On the dual nature of learning—a reinterpretation of "conditioning" and "problem solving." *Harv. educ. Rev.*, 1947, **17**, 102–148.
23. Mowrer, O. H. Comment on Estes' study: "Generalization of secondary reinforcement from the primary drive." *J. comp. physiol. Psychol.*, 1950, **43**, 148–151.
24. Mowrer, O. H. *Learning theory and personality dynamics.* New York: Ronald, 1950.
25. Murphy, G. *Personality: a biosocial approach to origins and structure.* New York: Harper, 1947.
26. Myers, J. A. An experimental study of the reinforcing value of food for non-hungry rats. M.A. Thesis, State Univer. of Iowa, 1949.
27. Saltzman, I., & Koch, S. The effect of low intensities of hunger on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*, 1948, **38**, 347–370.
28. Seward, J. P., & Handlon, J. H. The effect of satiation on the use of habit. *J. genet. Psychol.*, 1952, **81**, 259–272.

29. SEWARD, J. P., & LEVY, N. Maze learning by satiated rats as a function of secondary reinforcement. *Amer. Psychologist*, 1951, 6, 277–278. (Abstract)

30. SEWARD, J. P., LEVY, N., & HANDLON, J. H. Choice behavior in satiated rats as a function of drive during training. *J. genet. Psychol.*, 1952, in press.

31. SPENCE, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, 54, 1–8.

32. TEEL, K., & WEBB, W. B. Response evocation on satiated trials in the T-maze. *J. exp. Psychol.*, 1951, 41, 148–152.

33. TOLMAN, E. C. A drive-conversion diagram. *Psychol. Rev.*, 1943, 50, 503–513.

34. TOLMAN, E. C. There is more than one kind of learning. *Psychol. Rev.*, 1949, 56, 144–155.

35. TOLMAN, E. C. *Collected papers in psychology.* Berkeley: Univer. Calif. Press, 1951.

36. WEBB, W. B. The motivational aspect of an irrelevant drive in the behavior of the white rat. *J. exp. Psychol.*, 1949, 39, 1–14.

[MS. received February 14, 1952]

Psychological Review
Vol. 60, No. 2, 1953

LEARNING THEORY AND "ABNORMAL FIXATIONS"

JOSEPH WOLPE

University of the Witwatersrand, Johannesburg

In a recent article (5), Maier and Ellen, supplementing Maier's original thesis (4), have attempted to show that the persistent behavior induced in animals by certain conditions of stress cannot be understood in terms of learning concepts and must therefore be regarded as having a nonlearning origin which Maier has labeled "abnormal fixation" and which he believes to be a result of continued frustration often exacerbated by punishment. (The implication is that this behavior has as a basis a special kind of organic change qualitatively different from the neural modifications presumed to subserve learning (18).)

Maier and Ellen present their case in two parts. In the first they argue that certain features of the experimental results obtained by other workers are not satisfactorily covered by Mowrer's anxiety-reduction theory (13, 14); and in the second, that certain of the experimental findings in Maier's laboratories are inexplicable in terms of learning theory. In both cases they hold out frustration theory as the satisfactory alternative. The aim of the present paper is to show that learning theory is quite able to explain all the experimental facts, provided that it is remembered that there are reducible drives other than learned anxiety.

Comments on Some Data Obtained by Anxiety-Reduction Theorists

Maier and Ellen agree that in experiments such as those of May (10), Miller (11), and Mowrer (14) there is a learning of avoidance responses on a basis of anxiety reduction, but dispute the adequacy of this mechanism to explain all aspects of the experiments reported by Farber (2), Mowrer (15), and Mowrer and Viek (16).

Their quarrel with Farber rests entirely on the allegation that he does not account for the outstandingly great difficulty some of his rats had in extinguishing their anxiety-based responses. Yet Farber went to considerable pains to show that these animals came from strains that were more emotional generally. This seems a reasonable explanation, for higher emotionality would imply that stronger learnable anxiety responses would be evoked by the shock, and these would also be more strongly reinforced because higher measures of anxiety drive would be reduced. The effect of this would be that during the extinction series extinction would be counteracted by the reduction of unusually great learned anxiety drive.

In Mowrer's experiment (15) a rat was conditioned to run from the left end of a 4-foot alley to a safe compartment at the right end by consistently electrifying the floor of the alley 10 seconds after the introduction of the animal. When the habit of running within the 10 seconds' grace was well established, the *right half* of the floor was *permanently* charged. It was then soon found that the animal placed at the left end ran even more promptly to the safe compartment, and this behavior persisted for hundreds of trials.

Maier and Ellen reject Mowrer's "two-factor learning theory" explanation of this experiment, rightly, it seems, for one of Mowrer's main arguments in support of his theory has recently been shown fallacious by Miller (12) on the basis of an experiment by Sullivan (17). But this does not mean that we must fall

TABLE I

NUMBER OF RESPONSE INHIBITIONS OBTAINED IN SHOCK-CONTROL GROUP AND SHOCK-UNCONTROLLED GROUP

Successive days	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
Shock-control group	0	0	1	0	2	2	2	3	1	1	1	1	1	1	0	16
Shock-uncontrolled group	0	1	4	5	7	6	7	4	7	7	8	8	8	8	7	85

Cited from O. H. Mowrer and P. Viek (16, p. 396).

back on frustration theory, because, in fact, the results of the experiment are quite understandable in terms of established principles of learning, as follows.

During the early part of the experiment, emotional and running responses are reinforced to alley cues by reduction of the shock-induced drive when the animal reaches the safety compartment. Whenever the animal makes a run without being shocked, this constellation of responses is reinforced by anxiety-drive reduction. Now, when the right half of the alley is permanently charged, the ongoing responses of the running rat receive stronger reinforcement to the cues of that part of the alley, because when the safe compartment is reached, there is also reduction of the added measure of drive provided by the shock. Since these cues consist of wall and floor stimuli indistinguishable from those in the left half of the alley, it is not surprising that the animal's response of running when placed in the left half is reinforced and kept at a maximum by consistent shocks in the right half. This interpretation might be decisively tested by repeating the experiment with one half of the alley black and the other half white.

In Mowrer and Viek's experiment, (16) performed on two groups of 10 rats, the hungry rat was placed on the electrifiable floor of a rectangular cage, and offered food on a stick for 10 seconds. Whether the animal ate or not, shock was applied 10 seconds later. In the case of one group of ten rats, jumping into the air resulted in the experimenter switching off the shock (shock-controllable group). Each animal in this group had an experimental "twin" to which the shock was applied for the same length of time as it had taken its counterpart to jump into the air (shock-uncontrollable group). One trial a day was given to each animal. The number of animals in each group whose eating response during the 10 seconds was inhibited was charted each day with results shown in Table 1.

Mowrer and Viek try to explain the much smaller development of eating inhibitions in the shock-controllable group on the hypothesis that these rats have learned a "symbol for leaping" that somewhat alleviates their fear both before and during the shock. "By thus 'thinking' how the shock can be gotten rid of, the rat can presumably lessen the attendant fear of shock" (16, p. 198).

It is evident that *something* must be happening to check the development of fear responses to the stimulus situation that precedes the shock, but there is good reason for denying that the relevant occurrence is the evocation by this stimulus situation of a "thought" of the drive-reducing jumping response. The objection is that although states of drive can, through learning, become evocable by new cues, there is no evidence that this ever happens to drive *reductions*. *Thinking* how lunch may be obtained in

no wise appeases one's hunger. Cues along the learned pathway to the reduction of a drive do not reduce that drive. (See the writer's criticism (21) of Hull's use of "fractional goal responses" (3, p. 100).)

However, learning theory *can* explain how the reinforcement of the jumping response in the shock-controllable group could have brought about a diminishing level of anxiety in these animals. When an animal is subjected to continuous electric shock for the first time in a given situation, autonomic responses (anxiety) and numerous musculo-skeletal responses are evoked. A similar variety of musculo-skeletal responses will accompany the autonomic responses at every repetition of the electric shock unless there is one musculo-skeletal response that is repeatedly followed by cessation of the shock. If there is a response thus repeatedly reinforced, it will, of course, become increasingly dominant over all the other "competing" response tendencies. This implies a gradual weakening of the latter, probably by conditioned inhibition based on reciprocal inhibition (20). It is not unreasonable to suppose that *the weakening of "competing" response tendencies extends also to the autonomic responses*. Thus, in Mowrer and Viek's shock-controllable group, the repeatedly reinforced jumping response might have been expected to inhibit reciprocally the anxiety responses to some extent, thus slowly weakening their habit strength even though reduction of the shock-induced drive state would tend to reinforce both the jumping and the anxiety responses.

It appears from both experimental (19, 20) and clinical (22) observations that for a response to be weakened on the basis of reciprocal inhibition *the simultaneous competing response must be the dominant of the two*. This implies that the latter must be more "massive," i.e., subserved by a greater measure of central neural excitation (18). The jumping response reinforced by Mowrer and Viek was obviously rather massive; and it is conjectured that if they had chosen to reinforce, say, head turning to the left instead of jumping, the shock-controllable group would have developed much more anxiety than it did. Tending to support this conjecture is the fact that in the typical Cornell technique for producing experimental neuroses (e.g., 1) a very localized musculo-skeletal conditioned response comes to be increasingly dominated by autonomic anxiety responses (by a process discussed elsewhere (19)).

COMMENTS ON "ABNORMAL FIXATION" EXPERIMENTS

In Maier's laboratories rats were trained to jump to differential cards in a Lashley jumping apparatus. Jumping to the unlatched "correct" card was unrewarded with food, and to the unlatched "incorrect" card punished by a bump on the nose and a fall on to the net below. If, after the animal had learned a jumping habit, the cards began to be latched in random fashion, the animal would soon refuse to jump. It would then be forced to do so, usually by blowing a blast of air on to it. The jumping so compelled followed a pattern that would persist unchanged for hundreds of trials. Maier (4) called this "abnormal behavior fixation."

Maier and Ellen explore whether or not a learning-theory hypothesis will explain this phenomenon. They suppose that the air blast could result in anxiety becoming conditioned to the experimental situation. Then, jumping in a given way becomes persistent because, despite other changes in the situation, it continues to reduce the conditioned anxiety. Viewed in this way, ". . . an abnormal fixation would be considered as a learned adjustive mechanism." Maier and Ellen state that "on the surface"

this hypothesis seems satisfactory, and then proceed to point out a number of experimental facts that it fails to explain.

There is little doubt that the experimental facts they give do invalidate the above learning-theory hypothesis. But, as will be seen below, the same facts are quite in keeping with the differently formulated hypothesis that follows.

Each time a jump is forced by the air blast, it is reinforced by reduction of the air-blast-induced drive. This is *a primary drive,* and its reduction is clearly overwhelmingly responsible for the reinforcement; for the platform situation is at no stage conditioned to a sufficient level of secondary drive to be able to impel jumping in the absence of the air blast. When jumping in a particular way has thus been repeatedly reinforced, it becomes firmly established as the habitual response to the air-blast stimulus, and the more firmly it is established the weaker does the competing alternative response tendency become.

Armed with this interpretation we are in a position to deal with all of Maier and Ellen's objections, and we shall do so in the order followed by them.

1. If, for a fixated animal, one card (positive) always leads to food and the other (negative) to nose bumping, the strength of air blast needed to cause jumping gradually comes to be less when the positive card is on the fixated side than when the negative card is there (7). Maier and Ellen say that as this shows that the animal has learned to discriminate between the two cards, it should now jump to the positive card even when it is on the nonfixated side, and as it does not do so the fixation cannot be based on learning. But they overlook the fact that *the stimulus to jumping is the air blast and not the card.* The air blast inevitably evokes the one response that has been reinforced to it. Conditioned response tendencies to other stimuli in the situation can only be expected to modify quantitatively the conditioned jumping response.

2. Maier and Klee (8) found that if a jumping response to one side has become fixated, and an attempt is made to alter this behavior by latching the card on that side so that the rat is punished by bumping its nose, fewer changes occur if the animal is punished at every jump than if it is punished on only 50 per cent of trials, receiving food reward on the platform beyond the unlatched card on the other 50 per cent. Maier and Ellen correctly point out that this result cannot be explained in terms of the anxiety drive associated with the punishment. But it can be explained in terms of the drive due to the air blast. As mentioned above, repeated punishment increases the strength of air blast required to produce a jump towards a given card, while repeated food reward decreases it (7), and the mean air-blast requirement rises rather gradually with 50 per cent punishment. With 100 per cent punishment it would rise much more steeply. Increased air blast means increased drive with consequent greater drive reduction. Thus, with 100 per cent punishment the habitual response is more rapidly strengthened, and the chances of a jump to the other side rapidily decrease. That is why if an animal placed on 100 per cent punishment does not make the alternative response relatively soon, it does not make it at all. On the other hand, once the alternative response occurs, the ensuing air-blast-drive reduction and hunger-drive reduction immediately increase the probability of its occurrence on the next trial.

3. The bimodal distribution (7) is explained by the considerations of the previous paragraph. Only while the habit strength of the alternative response is strong enough for oscillation (2, 304–321) to permit it to occur occasionally,

is it possible for the fixation to be broken. A qualitative difference of mechanism is not implied by the bimodal distribution. In some animals the alternative response simply does not chance to occur while it may.

4. The persistence of the fixated response even though food is displayed in the other window (4) is not surprising. The animal automatically makes its strongly reinforced response to the powerful air-blast stimulus because it is in that way that neural organization has been modified by the learning process. If food-stimuli do not deflect the fixated response, it is because they command a relatively trivial habit strength in this situation.

5. Trial-and-error behavior does not occur when a fixated response is punished, because, in general, trial-and-error behavior does not occur when there is a well-learned response to the stimulus concerned (in this case, air blast).

6. The effectiveness of guidance in breaking a fixation seems to be entirely due to the fact that it makes the alternative response possible, taking the place of the no-longer-effective learned tendency.

7. Fixations are more rigid than most other habits because every evocation is followed by reduction of a primary drive.

Thus, the difficulties referred to by Maier and Ellen are all explicable by learning theory. Two findings from Maier's laboratories give incidental support to a learning-theory hypothesis. (*a*) With repeated punishment the strength of fixations increases between the 8th and 16th days of experimentation (6) in a way that might be expected of a habit. (*b*) With guidance, the position responses of animals are abandoned in much the same way whether they have had frustration training or reward training (9).

Of course, if the learning-theory explanation is correct, there is nothing abnormal about the Maier fixations. But crucial experiments are needed to clear up the issue finally. One that is suggested is to give rats the usual preliminary training and then force them by air blast to jump to cards that are always unlatched and never followed by food. Learning theory would predict that this would produce fixations, while frustration theory would presumably not expect this.

Meanwhile, in another communication (19), the writer has felt justified in the view that Maier's fixation experiments do not yield examples of experimental neurosis.

REFERENCES

1. ANDERSON, O. D., & PARMENTER, R. A long-term study of the experimental neurosis in the sheep and dog. *Psychosom. Med. Monogs.*, 1941, 2, Nos. 3 & 4.
2. FARBER, I. E. Response fixation under anxiety and nonanxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111–131.
3. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. MAIER, N. R. F. *Frustration—the study of behavior without a goal.* New York: McGraw-Hill, 1949.
5. MAIER, N. R. F., & ELLEN, P. Can the anxiety-reduction theory explain abnormal fixations? *Psychol. Rev.*, 1951, **58**, 435–445.
6. MAIER, N. R. F., & FELDMAN, R. S. Studies of abnormal behavior in the rat. XXII. Strength of fixation and duration of punishment. *J. comp. physiol. Psychol.*, 1948, **41**, 348–363.
7. MAIER, N. R. F., GLASER, N. M., & KLEE, J. B. Studies of abnormal behavior in the rat. III. The development of behavior fixation through frustration. *J. exp. Psychol.*, 1940, **26**, 521–546.
8. MAIER, N. R. F., & KLEE, J. B. Studies of abnormal behavior in the rat. XII. The pattern of punishment and its relation to abnormal fixations. *J. exp. Psychol.*, 1943, **32**, 377–398.
9. MAIER, N. R. F., & KLEE, J. B. Studies of abnormal behavior in the rat. XVII. Guidance vs. trial-and-error in the alteration of habits and fixations. *J. Psychol.*, 1945, **19**, 133–163.

10. MAY, M. Experimentally acquired drives. *J. exp. Psychol.*, 1948, 38, 66–77.
11. MILLER, N. E. Studies of fear as an acquired drive. I. Fear as motivation and fear reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, 38, 89–101.
12. MILLER, N. E. Comments on multiple process conceptions of learning. *Psychol. Rev.*, 1951, 58, 375–381.
13. MOWRER, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, 46, 553–565.
14. MOWRER, O. H. Anxiety-reduction and learning. *J. exp. Psychol.*, 1940, 27, 497–516.
15. MOWRER, O. H. Learning theory and the neurotic paradox. *Amer. J. Orthopsychiat.*, 1948, 18, 571–609.
16. MOWRER, O. H., & VIEK, P. An experimental analogue of fear from a sense of helplessness. *J. abnorm. soc. Psychol.*, 1948, 43, 193–200.
17. SULLIVAN, J. J. Some factors affecting the conditioning of the galvanic skin response. Unpublished Ph.D. dissertation. State Univer. of Iowa, 1950. Quoted by Miller (11).
18. WOLPE, J. Need-reduction, drive-reduction, and reinforcement; a neurophysiological view. *Psychol. Rev.*, 1950, 57, 19–26.
19. WOLPE, J. Experimental neuroses as learned behavior. *Brit. J. Psychol.* (General Section), 1953, 43, 243–268.
20. WOLPE, J. The formation of negative habits; a neurophysiological view. *Psychol. Rev.*, 1952, 59, 290–299.
21. WOLPE, J. The neurophysiology of learning and delayed reward learning. *Psychol. Rev.*, 1952, 59, 192–199.
22. WOLPE, J. Objective psychotherapy of the neuroses. *S. African med. J.*, 1952, 26, 825–829.

[MS. received March 10, 1952]

Psychological Review
Vol. 60, No. 2, 1953

TOLMAN'S INTERPRETATION OF VICARIOUS TRIAL AND ERROR[1]

GEORGE A. AUSTIN[2]

Laboratory of Social Relations, Harvard University

Taylor and Reichlin (5) have recently raised three basic objections to Tolman's interpretation of vicarious trial and error (VTE) behavior. They contend: (*a*) that his sowbug schema is unrelated to his general behavior theory and is relevant only to VTE; (*b*) that Tolman's theory of VTE does not conform to the facts; (*c*) that VTE is not an aid to learning. The aim of the present article is to show that the first two contentions are false and that Tolman is therefore justified in asserting that VTE aids learning.

Relation to Other Concepts

Tolman's sowbug schema (9, 11) employs two major concepts, "progression-readiness," which propels the bug forward and backward, and "orientation-readiness," which induces the bug merely to point at stimulus objects.

As Taylor and Reichlin remark, "progression-readiness" is a complex concept. It has components referring to association, perception, and need. We first define *hypothesis* (or instrumental belief) as a three-termed association: "If stimulus S_1 is apprehended (or registered, 13) and if response of type R_1 is made, then stimulus S_2 will be apprehended." An hypothesis becomes an *expectancy* if S_1 is in fact apprehended: "S_1 is apprehended, therefore if R_1 then S_2." An hypothesis is more strongly aroused into its corresponding expectancy, the more certainly the organism identifies a present apprehended stimulus as S_1. Furthermore, the more frequently, recently, and vividly the organism has apprehended stimulus S_1, the more certainly will he identify it *as* S_1. Expectancy is the cognitive immanent determinant of behavior (7).

To come to the purposive immanent determinant of behavior, we first define *cathexis* as a two-termed association between need and goal object (7, 13): "If need N (or drive D) is apprehended then the organism will 'persist through trial-and-error' and exhibit docility in thus persisting, to get to or from a goal object of type G." A cathexis becomes a *demand* or *motive*, the purposive component of behavior, when N is in fact apprehended. The stronger an apprehended need (the sowbug's progression-tension) the greater the demand. The reader will note the correspondence between hypothesis and cathexis, and between expectancy and demand.

Behavior results from the interplay of expectancy and demand; if stimulus S_2 of the expectancy is, or is associated with, goal G of the demand, then response R_1 of the expectancy will occur and the rat will no longer remain buried in thought. A given progression-readiness is composed of hypothesis and demand. Sample hypotheses are left card—jump—food and (after learning) white card—jump—food and black card—jump—shock. Tolman assumes that total behavior is determined by the algebraic difference between the progression-readiness for the stimulus it is now facing, and all other progression-readinesses, so that the progression vector (resultant progression-readiness) to jump to the rewarded white card is

[1] The author is indebted to Dr. J. S. Bruner and to Dr. E. C. Tolman for a critical reading of the manuscript.

[2] SSRC Research Training Fellow.

greater if jump to the black card has been punished. Any factor reducing the resultant forward progression vector tends to permit more VTE.

"Orientation-readiness" is less easily reducible to concepts Tolman has used in other contexts, principally because he has never systematically treated the "need" for new stimuli, curiosity, and exploration for its own sake. It would appear, however, that the properties of orientation-readiness are those of the need for novel stimulation, as the latter have been described in various accounts by Lashley, Maier, Hebb, Tolman himself, and others. The present discussion is confined to vision. The properties common to the two concepts are these: (*a*) Selectivity. The organism responds principally along one stimulus dimension (or to a particular stimulus aspect) at any given time. The dimension principally responded to depends on the stimulus and on innate and acquired characteristics of the organism, and may change during a sequence of stimulus presentations. (*b*) Best View. The organism will adjust its position and orientation with respect to a stimulus object in such a way that "satiation" or "need reduction" occurs most rapidly. This almost always means bringing the object to the center of the visual field by facing it. The organism may also approach or, more rarely, retreat from the object. (*c*) Discriminability. If one of two stimulus objects remains constant, an increase in (say, positive) physical difference between it and a second object is accompanied by a non-decrease, generally an increase, in discriminability on the corresponding psychological (behavioral) dimension (s); in other words, two objects are easy to discriminate if they are physically quite different. The more discriminably different another stimulus in the field from the stimulus at present attended to or "centered," the more readily will the organism shift its attention to, seek out, or (if the organism is a sowbug) point at the other stimulus. (*d*) Need Reduction. Attending to, or centering, a new stimulus reduces almost instantly the exploratory need for the aspect of the stimulus attended to. The more frequently and recently a particular aspect of a stimulus has been apprehended, the less in amount and the faster in time is the "need reduction" it produces, because the need is already at a reduced level.

Tolman has made two special assumptions for the sowbug, both related to Best View property (*b*). (*b′*) The nearer a peripheral stimulus to the center of the sowbug's perceptual field, the greater the tendency to point at the object and center it. This assumption is confirmed by the fact that VTE frequency is higher for smaller spatial angles between stimulus objects (10). (*b″*) In short-jump situations the approach tendency due to stimulus need is negligible. This assumption may indeed be *ad hoc*, but it is not implausible and so far as the author is aware, it is not contrary to fact or inconsistent with the body of Tolman's system.

In the presence of two stimulus objects to which it is responding discriminably, the sowbug will, as a result of properties (*b*) Best View, (*c*) Discriminability, and (*d*) Need Reduction, simply oscillate, facing one stimulus then the other, until the progression vector moves it forward in the direction it is facing.

On the whole, we may conclude that the schematic sowbug lies close to the center of Tolman's theory. Indeed, "behavior-adjustment," the process underlying VTE, was an early construct of Tolman's (6), and was designated as one of the three classes of behavior determinants, along with capacity and the immanent determinants (cognition and purpose) (7). Contrary to Taylor and

Reichlin, therefore, the sowbug's characteristics are not only consistent with Tolman's system but form an integral part of it. The sowbug is merely a means for focusing Tolman's theory on a particular area of behavior, or as Taylor and Reichlin remark (5, p. 389), it is an expository device. It follows that in deriving the facts of VTE behavior we may employ concepts and principles from any part of Tolman's writings.

DERIVATION OF VTE BEHAVIOR

Taylor and Reichlin's second major contention is that Tolman's theory does not account for certain facts of VTE behavior. They concede that it accounts for two facts, that VTE is more frequent for rats *learning* an *easy* discrimination (fact 2 in their list), and that VTE is more frequent for humans making a difficult discrimination (fact 3). Fact 2 is explained largely by Discriminability, property (*c*), which predicts a greater tendency to oscillate with more discriminably different stimuli, and the fact that at the learning stage of the task the organism has no strong hypotheses (i.e., he is "picking up his instructions," 12) and therefore has weak progression-readiness. Fact 3 is explained by the small difference between progression-readinesses related to the two stimuli; the small resultant progression-readiness permits more oscillation, more VTE. Both progression-readinesses are small for humans in part because of low demand due to weak need-tension. There is a more important element, however, not mentioned by Taylor and Reichlin, which must be introduced to account for the same fact observed in rats (16). This element is certainty of apprehension, of "making sure of which stimulus is which" (12). If the stimulus objects are hard to discriminate, not only may the subject's hypotheses be weak but these hypotheses will be less strongly aroused into their corresponding expectancies upon apprehension of the stimuli.

Fact 1 on Taylor and Reichlin's list refers to the low frequency of VTE movements per trial in the early trials, an increase in frequency which reaches a maximum when the subject has almost stopped making incorrect choices, and a subsequent decrease in VTE frequency. Taylor and Reichlin state (5, p. 391): ". . . Tolman's theory leads to the deduction that the number of VTE movements is a decreasing function of the number of trials," in contradiction to the maximum of fact 1. Their deduction is correct, however, only if the "first" trial is taken to be the trial on which the subject begins to respond fully along the correct stimulus dimension.

The main object of Tolman's second sowbug article (11, esp. p. 379) is to show how active selectivity of perception, i.e., "discontinuity" in learning, fits into VTE behavior. As long as the subject holds an irrelevant hypothesis, say "left card—jump—food," he will not VTE in response to the brightness dimension. As he begins to "learn his instructions," to develop the general hypothesis that brightness matters, he will begin to VTE. When he has learned "what to do" and while his specific hypotheses pertaining to the individual cards are still weak, VTE frequency reaches a maximum. Then, as Taylor and Reichlin correctly state, VTE frequency decreases because of increase in the strength of specific hypotheses. VTE depends not on number of trials as such, but on the interplay of variables at different stages in learning. Tolman's schema clearly accounts for the maximum of fact 1, and Taylor and Reichlin's criticism must be rejected.

Fact 4 refers to greater frequency of VTE if the gap to be jumped is narrow (8½ inches) rather than wide (23½ inches). Tolman explains this fact (8)

by assuming that the rat cannot discriminate as well over the greater gap distance, that stimulus differences become less prominent, and therefore the rat VTEs less. Taylor and Reichlin accept the derivation but object to the *ad hoc* assumption. But fact 4 may itself be *ad hoc*, in the sense that it really depends on the rat's "myopia." If this property of the rat is not parsimonious, so much the worse for parsimony. The question is, is Tolman's assumption true or not?

Taylor and Reichlin state (5, p. 391): "Hebb (2) has presented evidence that distant visual stimuli are more effective than near ones in determining the rat's behavior." Now Hebb's evidence indicates that if rats are presented with conflicting visual cues, within jumping distance and at several feet (walls and ceiling), respectively, they tend to respond to the more distant cues, presumably because distant cues are in general less variable. The VTE behavior under discussion, however, involves cues which are not conflicting and which lie at distances of much more nearly the same order of magnitude. The fact that rats respond to stimuli several feet away does not indicate that they have the same sensitivity to stimuli 8½ inches and 23½ inches away. To the author's knowledge there is, however, no direct evidence on the question.

Tolman suggests in the same paper (8) that caution, or resistance to jumping, explains the higher frequency of VTE in jump, as opposed to non-jump, situations (cf. 4). Tolman and Brunswik (14) earlier interpret caution as preference for "indifferent" as opposed to "ambivalent" means-objects; here, jumping stand as opposed to cards. But since resistance to jumping is presumably greater for wider gaps, the rats would tend to hold back and VTE more, contrary to fact 4. The differential resistance for the two gaps must therefore be small with respect to the differential effect of gap width on discriminability and to other motivational factors.

Taylor and Reichlin's fact 5 is that there is higher VTE frequency for shock-right or shock-wrong than for no shock. In criticizing Tolman here they first misinterpret his premise and then make an incorrect derivation. Since the premise, regarding the role of VTE in learning, merits separate consideration (see below), I shall simply show how Tolman does account for fact 5. First, shock forces the subject to pay closer attention and to emphasize the difference between the two cards (7). This accounts in part for the more rapid learning in shock situations; the rat "learns his instructions" sooner. Second, it tends to make the subject more cautious (8). For shock-right, the subject is ambivalent toward the reward card and therefore holds back from both cards, meanwhile VTEing. For shock-wrong, the reluctance to be shocked for making a wrong choice will hold the subject back and VTE until one of the stimuli is identified with a fairly high degree of certainty or until response occurs for one of three other reasons specified by Tolman (11).

Further support of this interpretation comes from the fact (3) that after the subject has learned the discrimination and jumps only to the reward card, he no longer VTEs for shock-wrong (because he can now identify the cards correctly) but continues to VTE for shock-right (because he is still being punished for jumping correctly). Thus both faster learning and greater VTE in shock situations are accounted for.

We conclude that Taylor and Reichlin's objections have been refuted; Tolman's theory accounts systematically for the major facts of VTE behavior.

ROLE OF VTE IN LEARNING

Taylor and Reichlin challenge Tolman's proposal that VTE aids learning.

They state his position as follows (5, p. 391):

"... when the animal is facing in the right direction, the stimulus 'reminds' him, so to speak, of the previous satisfactory consequences of responding to it, and that has the effect of increasing the relevant progression-readiness. Similarly, facing in the wrong direction diminishes the other progression-readiness."

Nowhere does Tolman say this. He *does* say that VTE indicates an "active selecting and comparing of stimuli" (12), which leads to better identification of the relevant stimuli and therefore to better learning (stronger hypotheses) and better performance (more accurate expectancies). Animals who learn quickly VTE more than do slow learners, and normal rats VTE more than brain-damaged rats (15). Both facts may be due to faster "satiation" from the apprehension of a stimulus object in smarter and normal rats. Except for one (apparently preliminary) experiment cited by Tolman (8), more VTE has always been reported concomitant with quicker learning and fewer errors for easy discriminations.

How can we define the real issue between Tolman, and Taylor and Reichlin? Let trial-and-error behavior be defined in brief as complete overt action resulting from expectancies having a low degree of certainty. Vicarious trial and error must occur, by definition, when the function of overt trial and error is accomplished by other means. The function of trial and error is central (the strengthening of hypotheses); the substitute, or vicarious, means may be partly central and partly overt, or completely central, if there is such a thing as thinking without motor accompaniment. Tolman uses the term VTE to refer to oscillating from side to side without forward motion. In this sense VTE could only occur in a discrimination situation.

Taylor and Reichlin, on the other hand, wish to refer only to "... incomplete or preparatory responses of the same type as the hesitations and false starts that may be seen when animals are learning to respond to a situation which does not call for a choice between alternatives" (5, p. 392). They interpret head movements to be not signs of vicarious trial and error in Tolman's sense, but signs of *incipient* [3] trial and error.[4]

The issue is whether VTE is predominately vicarious trial and error in Tolman's sense, or incipient trial and error in Taylor and Reichlin's sense. Let us note that Tolman's system does not exclude incipient trial and error as a mode of behavior. It is therefore maintained that Taylor and Reichlin's appeal to data gathered in a nondiscrimination situation, however material it may be in other connections, has little bearing on the present issue of VTE in discrimination situations. Since Tolman has accounted systematically for VTE results, it is incumbent upon Taylor and Reichlin to account for them more satisfactorily.

CONCLUSION

It has been shown that Tolman's sowbug schema forms an integral part of his theory of behavior; that the theory accounts for the facts of VTE, including others than those cited by Taylor and Reichlin; and consequently, that VTE may be interpreted as an aid to learning.

[3] A term suggested by Dennis and Russell (1).

[4] If Taylor and Reichlin are correct, "vicarious trial and error" is a misnomer. Perhaps we can best avoid theoretical bias by thinking of the letters "VTE" as standing for "vacillating trial and error," if we have to think of them as standing for words at all.

REFERENCES

1. DENNIS, W., & RUSSELL, R. W. Comments on recent studies of VTE. *J. genet. Psychol.*, 1939, **54**, 217–221.
2. HEBB, D. O. *Organization of behavior.* New York: Wiley, 1949.
3. MUENZINGER, K. F. Vicarious trial and error at a point of choice. I. A general survey of its relation to learning efficiency. *J. genet. Psychol.*, 1938, **53**, 75–86.
4. SCHLOSBERG, H., & SOLOMON, R. L. Latency of response in a choice discrimination. *J. exp. Psychol.*, 1943, **33**, 22–38.
5. TAYLOR, J. G., & REICHLIN, B. Vicarious trial and error. *Psychol. Rev.*, 1951, **58**, 389–402.
6. TOLMAN, E. C. A behaviorist's definition of consciousness. *Psychol. Rev.*, 1927, **34**, 433–439.
7. TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century, 1932.
8. TOLMAN, E. C. The determiners of behavior at a choice point. *Psychol. Rev.*, 1938, **45**, 1–41.
9. TOLMAN, E. C. Prediction of vicarious trial and error by means of the schematic sowbug. *Psychol. Rev.*, 1939, **46**, 318–336.
10. TOLMAN, E. C. Spatial angle and vicarious trial and error. *J. comp. Psychol.*, 1940, **30**, 129–135.
11. TOLMAN, E. C. Discrimination vs. learning and the schematic sowbug. *Psychol. Rev.*, 1941, **48**, 367–382.
12. TOLMAN, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, **55**, 189–208.
13. TOLMAN, E. C. The nature and functioning of wants. *Psychol. Rev.*, 1949, **56**, 357–369.
14. TOLMAN, E. C., & BRUNSWIK, E. The organism and the causal texture of the environment. *Psychol. Rev.*, 1935, **42**, 43–77.
15. TOLMAN, E. C., GEIER, F. M., & LEVIN, M. Individual differences in emotionality, hypothesis formation, vicarious trial and error, and visual discrimination learning in rats. *Comp. Psychol. Monogr.*, 1941, **17**, No. 3.
16. TOLMAN, E. C., & MINIUM, E. VTE in rats: Overlearning and difficulty of discrimination. *J. comp. Psychol.*, 1942, **34**, 301–306.

[MS. received March 25, 1952]

Psychological Review
Vol. 60, No. 2, 1953

SPENCE ON THE PROBLEM OF PATTERNING

M. E. BITTERMAN

University of Texas

In a recent issue of this JOURNAL (8) Professor Spence has provided a long-awaited extension of his fruitful theory of discriminative learning (7) to the problem of stimulus patterning. After having been concerned for many years with the implications of a purely summative formulation, he has now turned his attention to data which clearly fall beyond the scope of the earlier theory. The occasion for the new statement was provided by an experiment on the relative difficulty of simultaneous and successive problems of discrimination (11), in the analysis of which Spence was led to consider the question of how the successive solution is achieved. A problem which requires the animal to turn to the left at the choice point of a T maze when both alleys are dark and to turn right when both alleys are bright should, in terms of a summative theory, be impossible of solution since there is no differential reinforcement in any afferent dimension.

The new formulation does not discard, but instead builds upon, the earlier summative principle, and the result is a three-level theory of perceptual development. I. In the "standard situation" (simultaneous) "the excitatory strength of the positive cue . . . becomes steadily greater than that of the negative cue" (8, p. 89) as a consequence of differential reinforcement. II. The solution of the successive problem, in which each afferent component is equally often reinforced, is dealt with in terms of the concept of *compounding*; in the T maze described above the compounds *dark-left* and *bright-right* "acquire excitatory tendencies to the response of approaching" (8, p. 90) while the compounds *dark-right* and *bright-left* do not. Such problems are more difficult to learn, and compounds will develop only "when no one of the cue members is systematically reinforced more than the others" (8, p. 90). III. A still higher level of perceptual organization—*transverse patterning*—appears in problems which "require that the animal on each trial respond or not respond (by approaching) to a particular figure depending on what the other figure [is]. . . . According to the theoretical view proposed here, response on the basis of such cue-cue relations . . . would take place in nonarticulate organisms, but only under conditions that would not permit learning on the basis of a single reinforced component or some simpler type of cue-position pattern (e.g., the type of patterning involved in the successive problem)" (8, p. 92). Spence thus proposes a hierarchical conception of perceptual development; each level appears in the context of a specifically defined problem, and higher-level organizations develop only when the conditions for the development of lower-level organizations are not met.

While this extension of the earlier theory, which brings it into closer correspondence with the realities of perceptual organization, may be welcomed as a step in the right direction, considerable evidence is available to suggest that it has not been carried far enough. The principal weakness of the new formulation is to be found in the restricted role which is assigned to what Spence has called "transverse patterning." In the most general sense this term refers to the effect of variation in the grouping or configuration of stimuli in a multiple-

choice discriminative problem (e.g., the arrangement of cards in a two-window jumping apparatus) when the reinforcement of specific components or compounds is equated. This grouping effect is not limited to the rather specialized kind of problem which Spence describes, which suggests that what the older summative formulation requires is not supplementation but a more thoroughgoing revision.

Transverse effects in "simultaneous" problems. In a recent experiment Saldanha and Bitterman (6) trained rats in the jumping apparatus to discriminate concurrently between two pairs of cards—two black-and-white vertically striped cards differing in stripe width and two homogeneous gray cards differing in brightness. For Group A the positive stripe was paired with the negative stripe and the positive gray with the negative gray (Problem A), while for Group B the positive stripe was paired with the negative gray and the positive gray with the negative stripe (Problem B). Although the differential reinforcement of afferent components in this Level I problem was the same for both groups, a transverse effect showed up clearly. The animals of Group A mastered Problem A and then Problem B before the majority of animals in Group B mastered Problem B. These results were accounted for in terms of the opportunity which was afforded by Problem A for comparison of the stimuli to be discriminated. Similar results have been obtained by Coate (1) in the context of a continuity experiment.

Still a third experiment on the effect of pairing was conducted by Elam and Bitterman (2). Rats were trained in a jumping apparatus to discriminate between black-and-white striped cards differing in thickness under conditions which provided experience with both horizontally and vertically striped cards. When the thickness discrimination had been mastered, the animals were trained to discriminate between horizontal and vertical stripes irrespective of thickness. For one group the two thin stripes and the two thick stripes were paired, while for a second group the cards were paired in such a way that a difference in thickness was present along with the difference in orientation (i.e., thick-vertical with thin-horizontal and thin-vertical with thick-horizontal). The relational presentation of *irrelevant* components *retarded* discrimination.

In a second series of experiments by Teas and Bitterman (9) and by Turbeville, Calvin, and Bitterman (10) further evidence of a grouping effect in Level I problems was obtained. These studies were designed to investigate what may be called the *two-situational* problem. Suppose that animals learn concurrently a discrimination between two vertically striped cards differing in thickness and between two gray cards differing in brightness. The stripes and grays are presented as pairs and the lateral arrangement of each pair is varied systematically. The problem thus presents *four* card arrangements to the animals—the two lateral arrangements of each pair. In the corresponding two-situational problem each pair is presented in only one lateral arrangement—e.g., light gray on the left and dark gray on the right, thick stripes on the left and thin stripes on the right—with the positive card of one pair on the left and the positive card of the second pair on the right. In both problems, therefore, one of the grays is consistently reinforced while the other is consistently punished, and each spatial component is randomly reinforced and punished. According to both the earlier and the later versions of Spence's theory the two problems should be functionally equivalent. The results show, however, that the two-situational problem is significantly easier than the four-situational problem and that there is little transfer from the first problem to the second.

These results demonstrate that, even in the context of a problem which provides differential reinforcement of afferent components, different levels of perceptual organization may emerge depending on the grouping of the stimuli.

A third series of investigations by MacCaslin, Wodinsky, and Bitterman (5) on the process of stimulus generalization bear on the problem of transverse patterning. In one of these experiments a group of animals was trained to discriminate between horizontally and vertically striped cards (vertical positive) and a second group was trained to discriminate between the vertically striped card (positive) and a black card. The groups were equated for number of reinforcements on the positive card and then trained to discriminate it from another vertically striped card (negative) differing in thickness. The performance of the first group was superior to that of the second, demonstrating that the context in which reinforcement is given determines to an important extent the nature of perceptual development. An analysis in terms of components leads to precisely the opposite prediction on the basis of the inhibition which should be generated from the horizontally striped card of the first problem. Similar results were obtained in other experiments of this series.

Transverse effects in successive problems. An experiment by Wodinsky and Bitterman (12) demonstrates grouping effects in a problem which is possible of solution on the basis of cue-position compounds alone. Animals were trained in a three-window jumping apparatus on a problem involving black, white, and vertically striped cards. One group was reinforced for jumping to one window when three black cards appeared, to a second window when three white cards appeared, and to the third window when three striped cards appeared (e.g., *B*BB, W*W*W, SS*S*, where the positive cards are italicized). For the second group the card-position relations were identical except that the cards were grouped differently (e.g., *B*SW, S*W*B, WB*S*). As each animal in each group reached criterion on its problem, it was shifted to the problem of the other group. According to Spence's latest formulation, the two problems, which involve identical compounds, should be functionally equivalent; arrangement should make no difference because the problems can be solved on the basis of compounding alone. The results, however, show a clear transverse effect. The first problem proved to be more difficult than the second and transfer from each problem to the other was quite incomplete (initial responses were about 60% correct as compared with a chance level of 33%).

Theoretical and experimental constriction. Although these experiments clearly demonstrate the operation of transverse effects in first- and second-level problems, it may be anticipated that Spence will not consider them relevant to his theoretical position. For one thing, the experiments were performed with the jumping apparatus which involves the use of punishment for errors, a condition which Spence has "specifically avoided" (8, p. 89). Furthermore, in some of the experiments the correction method of training was utilized, although Spence has concentrated upon results obtained with the noncorrection method. Finally, there is reason to expect that Spence may regard the experimental designs as too complex. In discounting the results of Weise and Bitterman he remarks, for example, that "it is difficult to interpret the very complex type of discrimination set-up they employed. The simple discrimination situation is sufficiently difficult to deal with theoretically without adding all of the problems that arise as the result of the serial nature of the multiple-discrimination set-up along with the fact that it involves a gradient

of reinforcement" (8, p. 91).[1] Similar reasons might be advanced for discounting the experiments here described.

Now it is easy enough to protect a restricted theory by the design of restricted experiments, and it is easy to emphasize the limitations of such a theory by the design of broader experiments. The fundamental problem concerns the relative value of the two approaches. As Leeper (4) has noted, it is possible to take the position that research under restricted conditions will at least reveal the fundamental processes at work under those conditions, although the principles thus derived may have to be supplemented when the scope of research is broadened. Leeper himself believes that the principles derived from research under restricted conditions are likely to be superficial. "Such research," he maintains, "is not conducive to the discovery, in any deep and significant sense, of the fundamental processes at work even within a limited area" (4, p. 489). It might be well if this position were given serious consideration.

At other points in Spence's presentation restrictions seem to be brushed aside and his principles seem to acquire a more general flavor. This tendency is especially evident when he denies that "the type of patterning that the Gestalters and Gulliksen and Wolfle are talking about" may be evidenced in problems which can be mastered on the basis of differentially reinforced components or compounds (8, p. 92). It may be maintained on the basis of the evidence here reviewed that such effects do occur under certain experimental conditions. While the "complexity" variable is difficult to evaluate, the influence of the correction methods and punishment for errors should be studied in subsequent experiments. It may be anticipated that the results of such experiments will demonstrate the need for a more fundamental modification of Spence's theory than has yet been proposed.

[1] The experiment in question and the one by Spence provide the only comparisons of simultaneous and successive problems to be found in the literature. Spence should not have been "surprised" by the failure of Weise and Bitterman to consider Lawrence's data. Lawrence made no directly comparable studies of the two kinds of problem. Results identical with those of Weise and Bitterman (obtained with a modified jumping apparatus) will soon be published (M. E. Bitterman and Jerome Wodinsky, Simultaneous and successive discrimination, *Psychol. Rev.*, in press).

REFERENCES

1. COATE, W. B. Do simultaneous stimulus differences in the pre-training period aid discriminative learning? *Amer. Psychologist*, 1950, **5**, 257. (Abstract)
2. ELAM, C., & BITTERMAN, M. E. The effect of an irrelevant relation on discriminative learning. *Amer. J. Psychol.*, in press.
3. LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.
4. LEEPER, R. L. Review of Hull's *Essentials of behavior*. *Amer. J. Psychol.*, 1952, **65**, 478–491.
5. MACCASLIN, E. F., WODINSKY, J., & BITTERMAN, M. E. Stimulus-generalization as a function of prior training. *Amer. J. Psychol.*, 1952, **65**, 1–15.
6. SALDANHA, E. L., & BITTERMAN, M. E. Relational learning in the rat. *Amer. J. Psychol.*, 1951, **64**, 37–53.
7. SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
8. SPENCE, K. W. The nature of response in discriminative learning. *Psychol. Rev.*, 1952, **59**, 89–93.
9. TEAS, D. C., & BITTERMAN, M. E. Perceptual organization in the rat. *Psychol. Rev.*, 1952, **59**, 130–140.
10. TURBEVILLE, J. R., CALVIN, A. D., & BITTERMAN, M. E. Relational and configurational learning in the rat. *Amer. J. Psychol.*, 1952, **65**, 424–433.
11. WEISE, P., & BITTERMAN, M. E. Response-selection in discriminative learning. *Psychol. Rev.*, 1951, **58**, 185–195.
12. WODINSKY, J., & BITTERMAN, M. E. Compound and configuration in successive discrimination. *Amer. J. Psychol.*, 1952, **65**, 563–572.

[MS. received March 31, 1952]

Psychological Review
Vol. 60, No. 2, 1953

FRUSTRATION AND THE QUALITY OF PERFORMANCE: II. A THEORETICAL STATEMENT

IRVIN L. CHILD
Yale University
AND IAN K. WATERHOUSE
University of Melbourne

What is the effect of frustration on the quality of performance? There appears to be a dual tradition in the writings of psychologists and others who have given attention to this problem.

First, there is a tradition that frustration leads to improved quality of performance. Dewey's often cited account of why thinking occurs stresses the role of a problem or difficulty as the occasion for creative intellectual activity (8). Difficulty in such a situation often is an instance of frustration.[1] In more general accounts of the psychology of adjustment, unreduced tension is shown as giving rise to various forms of adjustment, of which some may be of high intellectual quality (25). On the level of society as a whole there are notions such as Toynbee's (26)—that the protracted existence of a challenge, often in the form of difficulty in meeting the needs of bare subsistence, is the condition for the joint constructive activity that produces a new civilization.

Second, there is also a tradition that frustration leads to lowered quality of performance. This is perhaps the more apparent part of the thesis of psychoanalysis and psychology of adjustment, since, on the whole, adjustments of poor quality to frustration have received the greater attention from therapists. This tradition is also evident in much of the discussion about the disorganizing effects of emotion (as reviewed, e.g., by Leeper [15]), inasmuch as emotion is often produced by frustration. Barker, Dembo, and Lewin's study of frustration and regression (2) is often cited in simple confirmation of this tradition, to the neglect of the rest of its content. Most recently this tradition is represented in Maier's systematization of the effects of frustration (20), as most of the effects he deals with would doubtless be considered to be of poor intellectual quality.

There is, then, an apparent conflict of belief in this matter. Indeed, the conflict appears strikingly in some general textbooks in psychology. In a chapter on thinking and reasoning frustration is viewed as the condition for more organized behavior, and in a chapter on emotion it is viewed as the condition for less organized behavior. The failure to use a common term such as frustration in the two chapters apparently permits the contradiction to go unnoticed.

Is this apparent contradiction due merely to failure to appreciate the role of severity of frustration, minor frustrations leading in fact to an improvement in quality of performance and major frustrations to the opposite, as might be inferred from the settings in which these contrary effects are often discussed?

[1] We are using *frustration* in a broad sense to refer to prevention of a person's direct progress toward a goal, not wishing to prejudge by definition the importance of various distinctions that can be made among the variety of events that fit this definition. We heartily agree with Brown and Farber's emphasis on the need to distinguish sharply between this definition of frustration and its definition as referring to a state of the organism (5, p. 480). But we feel it more useful to apply the term to the *event* of prevention of a person's progress toward his goal than to a *state* which may in some cases be inferred from the event.

Presumably not in any very uniform way, else why would anyone swear when he stubbed his toe, and how could any prisoner ever carry through successfully an ingenious plan for escape?

The greatest advance toward resolving this contradiction has been made by Barker (1) and by Barker, Dembo, and Lewin (2). By drawing upon their contributions, upon other aspects of psychological theory, and upon evidence obtained in a variety of pertinent studies, we hope to advance still further towards an understanding of the factors which influence the direction of change in quality of performance that results from frustration.

We have found it convenient to deal with three problems which it is useful to separate for purposes of analysis:

I. Effects of frustration in one activity upon the quality of performance in a second activity.

II. Effects of frustration in one activity upon the quality of performance in that activity.

III. Effects of frustration upon the quality of a person's behavior as a whole.

The three sections of this paper will be devoted to these three problems in turn. For the sake of brevity only one of these problems—the second one—has been selected for detailed treatment.

I. Effects of Frustration in One Activity upon the Quality of Performance in a Second Activity

The well-known experiment of Barker, Dembo, and Lewin is presented by those authors as dealing with a generalized effect of frustration upon the constructiveness of a person's behavior as a whole (2, p. 46). Actually, a critical analysis of the procedures and results indicates that it can only be said with certainty to deal with the effects of frustration in one activity upon the quality of performance in a second activity. The activity frustrated was children's play with a highly attractive set of toys; the second activity, in which quality of performance was measured, was play with a much less attractive set of toys. The theoretical discussion by Barker, Dembo, and Lewin, like their data, is most directly relevant to the problem of this section.

In discussing this problem Barker (1) and, less sharply, Barker, Dembo, and Lewin (2) make a definite contribution to an understanding of the factors which determine whether frustration will lead to a better or poorer quality of performance. The suggestion they make about frustration in relation to poorer quality of performance we would rephrase as follows: frustration of one activity will produce lowered quality of performance in a second activity to the extent that it leads to the making of responses which interfere with the responses of the second activity. Barker, Dembo, and Lewin minimize the role of this sort of hypothesis in explaining their results. We have shown in a previous paper (6), however, that their results actually support this hypothesis very strongly; and we feel that this is the most important empirical contribution of their study.

The opposite effect, improved quality of performance, is ascribed by Barker, Dembo, and Lewin to what we would call an increase, resulting from frustration of one activity, in the strength of drives which support the second activity. Barker (1) suggests three conditions under which such drives are likely to be strengthened in a way which results in increased quality of performance. We would rephrase them as follows:

1. When the second activity can be and is motivated in part by the original, unreduced drive which had been motivating the

frustrated activity, so that the second activity functions as a substitute for the first.

2. When frustration-produced drive leads to an attempt to escape from reminders of the frustrated activity, and preoccupation with the second activity is the mode of escape hit upon.

3. When the person was previously especially unmotivated with respect to the second activity, for it is then supposed that quality of performance may be favorably influenced by increased drive more than it is unfavorably influenced by interference.

These all seem to be significant suggestions, and in each case allied fields of research could provide evidence that indirectly supports their plausibility. They have not, however, been tested systematically in research on quality of performance, though they are drawn upon by Barker, Dembo, and Lewin in interpreting the behavior of individual subjects who in their experiment showed an increase instead of a decrease in constructiveness after frustration (2, pp. 179–186).

This contrast between the effects of interference and of increase in relevant drive, resulting from frustration, seems to us of fundamental importance, though it leaves many questions unanswered. This same contrast will be made in connection with the second problem, to be considered in the following section of this paper. Other points to be made in the following section can also be applied, with modification, to the present problem, but we shall discuss them explicitly only with reference to the second problem. There remains to be made here, however, a special point about the interference effect of frustration upon a second activity, a point which is distinctive for the problem of this section and essential for putting into proper perspective the role of frustration here.

The point is this: Frustration of the first activity may, *in comparison with active pursuit of the first activity,* actually increase the quality of the second activity by reducing the amount of interference with it. This is particularly likely to be true if the two activities are essentially alternatives of which the first activity is the preferred or dominant one. For if in this case the preferred activity is being pursued without frustration, all the overt responses which make it up are present to interfere with possible pursuit of the second activity. If, on the other hand, the preferred activity is thoroughly frustrated, there may remain, as possible sources of interference with the second activity, only implicit tendencies to return to the preferred activity. Interference arising solely from implicit tendencies, from thoughts, seems likely on the whole to be much less severe than interference arising from successful overt pursuit of a dominant activity. We suggest that one aspect of the Barker, Dembo, and Lewin experiment can probably be viewed in this light, though the design of their experiment does not permit our suggestion to be tested. We can only illustrate our meaning by suggesting a variation of conditions which was not actually used in their experiment.

The constructiveness of children's play with relatively unattractive toys was initially measured in a free-play period, with no other toys in sight. Later, the constructiveness of their play with these same toys was measured during a frustration period, in which the children had just been interrupted in play with more attractive toys and these more attractive toys remained in sight behind a wire barrier.[2] The constructiveness of

[2] For the purposes of the general point under discussion it should be noted that the play with the attractive toys is here regarded as the first activity, and the play with the unattractive toys is regarded as the second activity.

play with the unattractive toys was lower during the frustration period than it had been during the free-play period; but still, it was an activity of considerable constructiveness, or quality. Our contention is that the constructiveness of play with the attractive toys would not have been as high as it was, had it not been for the frustration arising from inability to play with the attractive toys. For, suppose that instead of being frustrated, the children had been allowed to continue play with the more attractive toys, the unattractive toys being put off by themselves in another part of the room. What, in this case, would have been the quality of performance in the second activity, i.e., interaction with the unattractive toys? We would predict that it would fall into a very much lower level still—that it would be largely confined to glances and sporadic beginnings of play, rapidly interrupted by return to the more attractive toys.[3]

Frustration of a preferred activity, then, may produce for a second activity a degree of interference which is intermediate—intermediate between the greater interference which would have occurred in the absence of the frustration and the lesser interference which would have occurred in the total absence of the preferred activity.

It is none the less true that frustration of one activity may constitute a genuine source of definite interference with a second activity. This fact seems to be clearly demonstrated in two experiments which were intended to deal with another problem, that of repression. In both cases there was, as it happened, no question of a more preferred and a less preferred activity; the instructions of the experimenter simply required the subject to work at two activities successively, and it was possible to observe the effect of frustration in one activity upon the quality of performance in the other activity. Sears (24) found that frustrating subjects in their attempts to do well at a card-sorting task led to a lowered quality of performance in a learning task. Zeller (27) found that frustrating subjects in their attempts to do well at the Knox Cube Test also led to a lowered quality of performance in a learning task. In both instances it appears that frustration in one activity led to internal responses (worry, for example) which interfered with maximally effective prosecution of a second activity.

It is probable that for Barker, Dembo, and Lewin's subjects as well, there was interference genuinely arising from the frustration. This is suggested by the fact that the overt responses of their subjects included attempts to escape from the situation, a response which doubtless contributed to the total interference and appears to be a response specifically to the frustration. But the total interference was probably much less than it would have been without the frustration.

In sum, then: Where quality of performance is lower than might be expected, and this lowering appears to be connected with the course of other activities, frustration of other activities is one possible source of interference; but successful pursuit of other activities may be a more important one. The college student who is frustrated in his attempts to arrange a date for the evening may not learn his German vocabulary that evening as well as he could; but it's a good bet that he'll learn it better than if he had had a date.

[3] This prediction, as applied to the Barker, Dembo, and Lewin experiment, is complicated by the fact that children could integrate the two sets of toys in a single play activity. For our point to be made, one must suppose that the rules of the situation did not permit this integration—a restriction which, for many situations to which one would wish to generalize, is imposed by the very nature of the activities.

II. EFFECTS OF FRUSTRATION IN ONE ACTIVITY UPON THE QUALITY OF PERFORMANCE IN THAT ACTIVITY

In dealing with the effects of frustration in one activity upon the quality of ongoing performance in that activity itself, we shall organize our discussion under five main headings. These represent five kinds of process or event which may influence the effect that frustration has on the quality of performance. This analysis has been difficult, because the several processes or events are closely interconnected and in many an instance would all be operating at once. We believe, however, that this sort of analysis is useful for reaching an understanding of the effects we are dealing with.

A. Extinction of the initial response to the situation

When a person is frustrated in some activity, the situation to which he is responding is thereby somewhat changed. The extent to which it is changed, however, varies, and in some instances it may be useful, in predicting his response, to consider the situation to which he is responding as essentially the same as it was before the frustration. Where this is a useful approach to make, Hull's concept of the habit-family hierarchy (12), expressed in a somewhat more general form, suggests the importance for our problem of the extinction of the initial response to the situation.

A person may be conceived of as having, in any specific situation, tendencies to make various response sequences which may all potentially lead to the goal towards which he is oriented. These various response sequences may be thought of as a hierarchy, the various members of which differ in habit strength (that is, in the strength of the tendency for them to be evoked). The sequence for which the habit strength is initially strongest will be the one first evoked. If the resulting activity is frustrated, its habit strength is diminished by the process termed extinction. With persisting frustration, its habit strength may be reduced below that of the other members of the hierarchy. At this point the other response sequences in the hierarchy will begin to be evoked. The effect of frustration upon the quality of performance in this case, then, will depend upon the relative quality of the initial response sequence which is extinguished and of the other response sequences which are then evoked instead.

On the whole, it seems likely that the initial response sequence will be the sequence of highest quality in the hierarchy. The reason for this expectation is that the response sequence of highest quality is likely to have been the most strongly and consistently rewarded in similar situations in the past, and thus to have become the response sequence of greatest habit strength and the one first to be evoked.[4]

Several experiments on frustration appear to illustrate this effect of extinction of initial response upon quality of performance, though in each case it is quite likely that other processes to be considered later were also influential. Three experiments can serve as particularly apt examples. In each case the subject was required to engage in some more or less intellectual task. In an experiment by Sears (24), and in one by McClelland and Apicella (16), the task involved card sorting; in an experiment by Postman and Bruner (22), the task consisted of attempting to perceive and report words as their exposure time was gradually increased from a subliminal value. In a subject faced with any of these tasks, there is evoked a general mode of responding which is likely to be

[4] We are indebted to Dr. Gregory A. Kimble for suggesting this point to us.

about the most adaptive of which he is capable. But in each of these experiments, as the subject responded and continued to respond in this adaptive manner, the experimenter withheld the normal reward of knowledge of satisfactory performance, and instead told the subject he was failing miserably, doing worse than anyone else, etc. The lowered quality of performance which then appeared in each of these experiments may well have been due primarily to the extinction, resulting from nonreward, of the subject's adaptive response tendencies, together with the fact that for this situation the subject did not have any alternative response tendencies which would at all rival the initially dominant one in the quality of performance to which they would lead.

The effect of extinction of the initially dominant response tendency is not, of course, necessarily a lowering of quality of performance. The order of response sequences in the hierarchy for the given situation may be determined by generalized effects of learning which took place in a previous situation (or situations) which was appreciably different from the present one. In particular, the previous and present situations may differ in the quality of performance which would be judged to characterize particular response sequences if evoked in those situations. Thus the present situation may first evoke a response sequence which was of high quality in former situations but of low quality in this situation; extinction of the tendency to respond with this sequence *may* in that case lead to the evocation of a response sequence lower in the hierarchy but of higher quality in this situation.

B. Situational changes

In the preceding section we considered certain implications that follow when the frustrated person may be considered to be still responding to essentially the same situation. In this section we turn to certain implications which follow when the character of the frustrating circumstances is such that the person must be considered to be now responding to a situation very different from that which preceded the frustration. We shall not deal here with the fact of frustration itself as a new element in the situation to which the person is distinctively responding; this matter we shall discuss in Section D. We shall deal here simply with specific changes in the situation which are inherent in the specific manner by which the frustration is brought about.

The point we are concerned with here is this: One effect of frustration is to alter the person's situation in such a way that behavioral possibilities are changed, and this alteration has implications for the possible quality of the person's performance.

On the one hand, frustration may alter the situation in such a way as to render impossible any responses of high quality directed at the original goal. There is an approach to this condition in the Barker, Dembo, and Lewin experiment (2). The highly constructive behavior of complex play with the desirable toys was rendered impossible by making those toys completely unavailable to the child. If the constructiveness of behavior in relation to the goal of playing with those inaccessible toys was characteristically reduced by frustration (no systematic evidence was in fact collected on this point), was it not largely because this highly constructive behavior was made impossible and no other equally constructive behavior in relation to that goal was possible for most of the children? Parallel examples from nonlaboratory situations come readily to mind. For the man whose beloved marries someone else, the formerly most constructive behavior of striving by appropriate means to gain

her affection is now impossible if he accepts the morality of this society. Indeed, the situation is now so changed that there may be no possibility at all of constructive behavior directed at his original goal; if he is still so strongly driven toward this goal as to make some kind of response in that direction, it must of necessity be of poor intellectual quality, as, for example, wish-fulfilling fantasies, or various nonconstructive social acts.

On the other hand, frustration may alter the situation in such a way as to make possible the achievement of the goal by acts of higher intellectual quality than were previously possible or appropriate. The man who is digging a hole with his spade, in order to plant a tree, has no more constructive behavior open to him than the simple routine of digging; though this is the most efficient and adaptive behavior under the circumstances, it is not of very high intellectual quality. But if the handle of the spade breaks, and he is thus frustrated, more complex and more constructive behavior, of higher intellectual quality, now becomes possible and indeed essential as a means to the original goal. The skill of digging with a spade must now be integrated with the skill of shaping a new handle, or with the social skills involved in borrowing or buying another spade in a much more complex sequence of behavior leading to the original goal. This sort of effect may be seen in the Barker, Dembo, and Lewin study if one looks solely at the means by which the child achieved, or might have achieved, contact with the desirable toys. When they were freely available to him, he simply approached and touched them. When a barrier was interposed, the only behavior that might possibly have led him to these toys was a much more complex sequence of influencing the experimenter, though, as it happened, it had been predetermined that even this should not be successful.

C. Quality of the responses available for performance

In sections A and B we have shown that the elimination of one response, as a result of frustration, may influence the quality of performance. In section A, we considered the elimination of one response through extinction. In section B, we considered the elimination of one response because of the removal of some kind of environmental facility or support which is essential for its performance. Just how this elimination of one response will affect the quality of performance depends, of course, both upon the quality of the eliminated response and upon the quality of the other responses which then come to be made. We must now consider explicitly, therefore, the question of what variables influence the quality of the responses available in the person's repertoire and likely to be made if frustration eliminates the initial response.

In the case we have dealt with in section A, where the person may be considered as responding to essentially the same situation before and after frustration, we have already suggested that Hull's concept of the habit-family hierarchy provides a useful theoretical schema for dealing with this problem. The quality of the new behavior resulting from frustration would be predicted from the quality, as responses in this situation, of the responses next to the initially dominant one in the hierarchy. Actual application of this schema, of course, requires measurements both of quality and of habit strength of response sequences. Such measurements are certainly possible for complex human behavior, and have been made in connection with other problems. With reference to studies already done which are directly relevant to this problem, how-

ever, the schema can only be applied by using gross judgments of great differences in quality and habit strength between the initial response sequence and other response sequences, as in our interpretation of the experiments we cited in Section A.

In the case we have dealt with in Section B, where the situation to which the person is responding must be considered as radically changed, the same theoretical schema of Hull's may be considered as sometimes applicable. Here the quality of performance after frustration would be predicted from knowledge of the quality represented by the response sequence *highest* in the habit-family hierarchy for this new situation. We know of no research studies, on just this point, to which this mode of analysis is readily applicable. It is obviously applicable, however, to incidents of everyday life. Imagine a person driving his car to work who is frustrated by a flat tire, which radically changes his immediate situation. The quality of his response, e.g., swearing and sulking vs. changing the tire or calling a repair man, seems likely to be influenced by what particular response tendencies to this changed situation have become dominant as a result of his previous experience in similar situations. This sort of analysis should also be useful in leading to systematic research.[5]

Regardless of how the elimination of the initial response is brought about, however, the concept of the habit-family hierarchy is not by itself adequate to deal with all cases. For in many cases the initial response sequence is replaced, not by some other response sequence which has a predictable habit strength resulting from previous reward or nonreward in similar situations, but rather by some novel response sequence which has never previously been performed by the person in any situation.

Now, under these circumstances, the problem of predicting the effect of frustration upon quality of performance becomes the specific problem of predicting whether a person, in the face of frustration, will produce novel responses and whether these novel responses will be of high or low intellectual quality. This specific problem is one to which a considerable body of scientific research is relevant. Relevant research is indeed so voluminous that we cannot hope to review it here. All we wish to do here is call attention to its relevance to the problem of frustration, for most of this research has been performed, and has been discussed, in contexts far afield from the dynamics of frustration reactions.

First of all, there is research on intelligence as an organismic variable which influences the person's reactions in a variety of situations. If intelligence tests measure so broadly relevant a variable as is often hoped, that variable should be highly useful in predicting the quality of a person's response to frustration—in predicting, in other words, the likelihood that a frustrated person will hit upon a novel response of high quality rather than persisting in an unsuccessful response or making novel responses of poor quality.

Second, if this first point is correct, research on determinants of intelligence is also relevant to the present problem. If heredity, nursery-school training, institutionalization, intellectual character of the home environment, etc., influence general intelligence, they should influence the likelihood that the frustrated person will make a novel response of high quality.

[5] The only close parallel in systematic research that we know of is in a study by Davitz (7), which is concerned with the problem of section I of this paper. He frustrated children in an enjoyable activity of watching movies and eating candy. Their subsequent responses in a free-play situation were found to be influenced, in a way relevant to quality of performance, by previous training in a similar situation.

Finally, research on factors in the immediate situation which influence the adequacy of reasoning and problem solution is relevant to the present problem. Such research has not ordinarily been formulated as dealing with frustration. When we speak of frustration, we ordinarily think of a person as at first anticipating steady progress toward his goal, and at a later point encountering a barrier. In experimental studies of reasoning and problem solution, on the other hand, the barrier is generally present at the outset; the subject is asked to orient himself toward a goal which is obviously difficult to attain. But the determinants of quality of response under these special circumstances should certainly help to illuminate also response to problem situations as they arise in the form of frustration in normal life.

As this research on situational variables in problem solution is less widely known than that on intelligence and its determinants, it may be pertinent here to cite examples. A series of experimental studies by Maier provide apt examples. Maier demonstrates that the likelihood that a subject will make the novel and correct response to a problem situation is influenced by demonstration to the subject of part-responses which are required for it (18), by giving to the subject a particular kind of part-response which Maier terms "direction" (17), and by instructions designed to establish a general set towards flexibility of response (19). Such variables as these seem clearly applicable to understanding the quality of response to frustration in everyday life.

Much of the research on the availability of novel responses of high quality has been done in a strictly empirical context. This research is too diversified for us to attempt here any theoretical integration. We would like to point out, however, that the same kind of behavioristic analysis which we apply elsewhere in this paper is applicable here too. Examples of its application may be found in a theoretical paper by Hull (13) which discusses an experiment by Maier on "reasoning" in rats, and in Dollard and Miller's recent discussion of problem-solving behavior in human beings (9). That such an application of systematic theory may be fruitful for research in this area may be illustrated by recent studies by Birch (4) and by Gladstone (10), which demonstrate an influence of learning on the availability of responses for problem solution.

D. Habits of responding to frustration

We have so far considered the person as responding to the situation in which frustration is occurring, but not as responding to the fact of frustration per se. But the occurrence of frustration is, of course, itself a distinguishable aspect of the situation to which the person may respond distinctively. A person might conceivably have general habits of responding to all frustrations, or he might have more specific habits of responding to particular classes of frustration which for him were distinctive. The possible responses which might, in various individuals, have come to be elicited by the cue of frustration, are of course innumerable. We propose to call attention here to several classes of response which appear to have a special relevance for the influence of frustration upon quality of performance.

1. *Persistence vs. withdrawal.* Persistence in striving for the goal, in the face of frustration, is a response which keeps the individual in the situation and makes possible the emergence of novel responses of high intellectual quality, though whether such responses do in fact emerge will then depend upon such variables as those considered in section C. The degree of persistence appears to be in part determined by habits of response to frustration. Grosslight and Child (11) showed, for one experimental situation, that subjects who had been sub-

jected to frustration in the experiment and rewarded for persistence, subsequently persisted much longer in the face of continuous frustration than did subjects who had experienced only success until the time of continuous frustration.[6] In the same study tentative evidence was found that the first group of subjects, as a result of their persistence, were more likely to make novel or creative responses of a sort which under many circumstances would lead to a removal of the frustration. The second group of subjects, on the other hand, were more likely simply to withdraw from the situation or confine their responses to mere staring.

2. *Interfering responses.* Another difference among persons in their habits of responding to frustration has to do with their tendency to make responses which interfere with effective pursuit of the original activity and thus lower its quality. Thus Waterhouse and Child[7] used a questionnaire to measure the extent to which individuals habitually respond to frustration with potentially disruptive reactions, such as aggression, self-blame, and self-justification. They found that people scoring high on this personality measure, when subjected to experimental frustration, showed a lowered quality of performance; people scoring low on this personality measure, on the other hand, when subjected to the same experimental frustration, actually showed an improved quality of performance. A closely parallel finding was obtained by Mandler and Sarason (21; cf. their Predictions 4 and 5) in a study using a different kind of intellectual performance and a different measure of tendency to make interfering responses, a questionnaire designed to measure degree of anxiety as a response to examination situations. This finding is still further confirmed in another experiment by Sarason, Mandler, and Craighill (23). We may tentatively conclude that strong habits of responding to frustration, or to the anticipation of frustration, with responses which interfere with a complex intellectual activity tend to lower the quality of such activity when it must be performed in a frustrating situation.

3. *Drive-producing responses.*[8] Among the responses a person may make to frustration are internal responses which create or strengthen drive states. Indeed, some of these drive-producing responses are among the interfering responses we have mentioned in the preceding paragraph. Drive-producing responses, however, have two special properties in relation to the present problem.

(*a*) Certain drive states produced in response to frustration may operate to increase the motivation supporting the goal-oriented activity and thereby to improve the quality of performance. Individuals who have habitual tendencies to react to frustration with responses which create or strengthen these particular drive states would then improve their performance in the face of frustration. There appears to be no published research which provides clear evidence directly relevant to this notion, or which would indicate what particular drive states operate in this way, but this gen-

[6] A finding, of course, parallel to the typical outcome of experiments on partial reinforcement using traditional conditioning techniques (14).

[7] Frustration and the quality of performance. III. An experimental study. *J. Pers.*, 1953, 22, in press.

[8] Brown and Farber (5) have recently published an article which, while not focused on the problem of quality of performance, is highly relevant at this point to our treatment of this problem. Their "'emotional' interpretation of frustration behavior" might be regarded in large part as a much more thorough attack on the problem we deal with under the label of "drive-producing responses." We differ from them in viewing an emotional interpretation, and what they call nonemotional interpretations, which would include most of the rest of our treatment, not as alternative approaches (5, p. 480) but as two aspects of theory which need to be put together for the prediction of behavior.

eral notion appears to us to be a useful guide to future research.

(*b*) Quality of performance is likely to be greatly influenced not only by the drive states created by frustration, but also indirectly by other responses which are evoked by those drive states. The individual's habits of responding to drive states—in particular to the drive states likely to be evoked by frustration—thus are crucial in determining the effect of frustration upon the quality of his performance in the original activity. Among the drive states likely to be evoked by frustration are states of intense general emotion. These emotional states provide an apt example to illustrate the point we wish to make here.

Psychology textbooks often refer to the disorganizing effects of severe emotion (15). Undoubtedly severe emotion does often have a disorganizing effect and thus reduces the quality of performance in the face of frustration. In part this may be because the emotional responses themselves are to some extent incompatible with the ongoing instrumental activity. But in even greater part the disorganizing effect may have to do with responses *to* the emotional state. A typical person in our society is likely to have well-established tendencies to react to strong emotion with various responses—such as withdrawal from the emotion-arousing situation, close attention to the emotional experience, worry, expressive behavior such as swearing and gesturing—which all tend to interfere with efficient pursuit of the original goal-oriented activity. We would suspect that persons with a different habit structure might react to the same emotional states in themselves with a higher, rather than a lower, quality of performance. This appears to be the assumption underlying certain aspects of military training and implicit in the belief that seasoned troops are more dependable than inexperienced ones—the assumption that training can modify the way a person responds to an intense emotional state, indeed can modify it so radically that intense emotion may come to have an organizing rather than a disorganizing effect on behavior.

E. *Situational and task variables in relation to the fact of frustration*

In section D we have shown that the person's habits help determine whether his response to the fact of frustration will be such as to improve, or such as to detract from, the quality of his performance. In this section we wish only to point out briefly that the person's response to the fact of frustration will also be influenced by a variety of situational and task variables. The same kinds of response to frustration remain pertinent here.

Differences in instructions or in initial set given by the situation, for example, may influence the likelihood that frustration will evoke persistent striving or, on the other hand, withdrawal. Various specific circumstances in the situation may help determine whether frustration evokes responses which interfere with the original activity, and what effect it has on drive states and on responses to these. The extent to which heightened drive can lead to improved performance, and the extent to which other responses are incompatible and produce interference, may vary with the exact nature of the task or activity in which quality of performance is being judged. In all these ways, then, the kinds of responses we have considered in section D are relevant to the effect of frustration on quality of performance, but relevant not only as a function of the person's habits but also as a function of situational and task variables.

III. EFFECTS OF FRUSTRATION UPON THE QUALITY OF A PERSON'S BEHAVIOR AS A WHOLE

In the analysis we have presented thus far, frustration and its consequences may

play a very unimportant part in the total life of the person. Yet there is nothing about the analysis we have presented that restricts it to such cases. We wish to illustrate here the applicability of the analysis to cases of more pervasive effects on quality of behavior, though detailed application to any one problem is beyond our present intent.

One sort of case in which the question of a pervasive effect of frustration on quality of performance arises is the situation where the frustrating circumstances are such as to interfere with not one, but a great many of the person's strivings. Several such situations have been studied by psychologists and other social scientists, e.g., internment in a concentration camp, unemployment, and subjection to severe acculturation pressures. As a single example, to illustrate the applicability of our analysis, we refer to Bettelheim's study of concentration camp inmates (3).

Where very widespread frustration is imposed by the environment, as in concentration camps, it seems clear that the elimination of old responses, partly through impossibility of performance and partly through extinction, is a first determining factor. But with old responses eliminated there is a wide range in the quality of the new responses that appear. The availability of specific responses to this situation will be one potent influence; an example is the role that Bettelheim's scientific background must have played in enabling him as an individual to make a highly constructive adjustment to a concentration camp, through reacting to the situation in part as an intellectual challenge (3, pp. 422–424). Another potent influence is likely to come from the individual's habits of responding to frustration; an example of this is provided in Bettelheim's account of the persistent submission to authority of the German middle class as a factor interfering with a constructive adjustment to a concentration camp (3, pp. 425–426). Finally, the role of situational factors may be illustrated indirectly by Bettelheim's point that the typical adjustment of various categories of prisoner differed because of the different significance of imprisonment for them in view of their backgrounds (3, pp. 424–429).

A somewhat different kind of case in which the question of pervasive effects on quality of performance arises, is the case in which a person's behavior is simply noticed to be conspicuously high or low in general quality, and there is the possibility that this condition is in part a product of psychodynamics. Psychotics, some neurotics, and some apparently feebleminded persons may be seen as persons with a generally low quality of response which seems to result from dynamic processes of adjustment to life situations. "Frustration" is certainly no adequate label—if there be one—for the great variety of situations to which is attributed an important role in these processes. Yet frustration certainly is one of the variables relevant to understanding these processes. The opposite extreme, of the genius with a generally very high quality of performance, has been little studied from a psychodynamic point of view, so that we can hardly venture a judgment about the probable role of frustration here.

In cases of neurosis or psychosis the applicability of our general approach is very aptly illustrated by certain aspects of the recent book by Dollard and Miller (9). In their interpretation of such pervasively maladjustive behavior, they stress the role of a person's previously established habits of response to strong emotion or to specific drives. In particular, habits which prevent the correct labeling of an emotion or drive tend to reduce the quality of behavior by eliminating the potentialities for fine discrim-

ination and appropriate generalization which may be brought into a behavior sequence by the use of correct labels. An account of neurotic or psychotic behavior in such terms tries to explain much more than the patient's reaction to frustration; it tries to explain also, for example, his reactions to goal attainment. But the point relevant here is that, if this is a useful explanation of neurotic behavior in general, it is also specifically an explanation of why neurotics may tend to make responses of poor quality in frustrating situations.

REFERENCES

1. BARKER, R. G. Frustration as an experimental problem. V. The effect of frustration upon cognitive ability. *Charact. and Pers.*, 1938, 7, 145–150.
2. BARKER, R. G., DEMBO, T., & LEWIN, K. Frustration and regression: an experiment with young children. *Univer. Ia. Stud. Child Welf.*, 1941, 18, 1–314.
3. BETTELHEIM, B. Individual and mass behavior in extreme situations. *J. abnorm. soc. Psychol.*, 1943, 38, 417–452.
4. BIRCH, H. G. The relation of previous experience to insightful problem-solving. *J. comp. Psychol.*, 1945, 38, 367–383.
5. BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, 48, 465–495.
6. CHILD, I. L., & WATERHOUSE, I. K. Frustration and the quality of performance. I. A critique of the Barker, Dembo, and Lewin experiment. *Psychol. Rev.*, 1952, 59, 351–362.
7. DAVITZ, J. R. The effects of previous training on postfrustration behavior. *J. abnorm. soc. Psychol.*, 1952, 47, 309–315.
8. DEWEY, J. *How we think.* Boston: Heath, 1910.
9. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.
10. GLADSTONE, A. The internalization of counting. Ph.D. dissertation, Yale Univer., 1949.
11. GROSSLIGHT, J. H., & CHILD, I. L. Persistence as a function of previous experience of failure followed by success. *Amer. J. Psychol.*, 1947, 60, 378–387.
12. HULL, C. L. The concept of the habit-family hierarchy and maze learning. *Psychol. Rev.*, 1934, 41, 33–52 and 134–152.
13. HULL, C. L. The mechanism of the assembly of behavior segments in novel combinations suitable for problem solution. *Psychol. Rev.*, 1935, 42, 219–245.
14. JENKINS, W. O., & STANLEY, J. C., JR. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, 47, 193–234.
15. LEEPER, R. W. A motivational theory of emotion to replace "emotion as disorganized response." *Psychol. Rev.*, 1948, 55, 5–21.
16. MCCLELLAND, D. C., & APICELLA, F. S. Reminiscence following experimentally induced failure. *J. exp. Psychol.*, 1947, 37, 159–169.
17. MAIER, N. R. F. Reasoning in humans. I. On direction. *J. comp. Psychol.*, 1930, 10, 115–143.
18. MAIER, N. R. F. Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *J. comp. Psychol.*, 1931, 12, 181–194.
19. MAIER, N. R. F. An aspect of human reasoning. *Brit. J. Psychol.*, 1933, 24, 144–155.
20. MAIER, N. R. F. *Frustration: the study of behavior without a goal.* New York: McGraw-Hill, 1949.
21. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, 47, 66–173.
22. POSTMAN, L., & BRUNER, J. S. Perception under stress. *Psychol. Rev.*, 1948, 55, 314–323.
23. SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, 47, 561–565.
24. SEARS, R. R. Initiation of the repression sequence by experienced failure. *J. exp. Psychol.*, 1937, 20, 570–580.
25. SHAFFER, L. F. *The psychology of adjustment.* Boston: Houghton Mifflin, 1936.
26. TOYNBEE, A. J. *A study of history* (abridgement of volumes I–VI by D. C. Somervell). New York: Oxford, 1947.
27. ZELLER, A. F. An experimental analogue of repression. II. The effect of individual failure and success on memory measured by relearning. *J. exp. Psychol.*, 1950, 40, 411–422.

[MS. received February 13, 1952]

AGNOSIA IN ANIMAL AND MAN [1]

JOSEPHINE SEMMES [2]

New York University—Bellevue Medical Center

To a psychologist from Mars, it might seem paradoxical that for us the study of integrative mechanisms of behavior has become almost exclusively the province of the animal psychologist. Complex processes, such as learning, problem solving, and concept formation, are studied primarily in infrahuman animals, which, whatever their virtues, are not noteworthy for these abilities. This displacement of the human subject by the rat or monkey appears to rest on two related articles of faith: First, that the taxonomy of cognition is mature enough to allow study of the interrelation of established variables in the simpler and better-controlled situations that the use of animals permits. Second, that the laws and mechanisms which are uncovered in one species will apply equally well to others, and that a complete "phylogeny of behavior" merely awaits appropriate adjustment of the constants in some fundamental equations.

No doubt these beliefs have considerable heuristic value, but since, whether we will or no, generalizations from animal studies will be made to man, the validity and adequacy of such extrapolations should be tested empirically. Perhaps there are certain questions, such as localization of function in the brain, which can never be answered for man with the same certainty as can be attained for animals. Yet, conversely, perhaps there are questions which are logically prior to anatomico-functional correlation, to which experimentation on man alone is likely to provide answers. Cognitive processes, as far as we know, reach their highest development in the human species. Language exists only in rudimentary form, if at all, in other animals. Because of his capacities, injury or disease may produce in man disorders such as the agnosias, apraxias, and aphasias, which offer an extraordinary opportunity for study of the structure of mind.

The opportunities presented by lesion of the nervous system in man are not now being exploited by any large or organized group of psychologists. Most physiological psychologists scorn clinical material, because exact knowledge of the locus of lesion in human cases is usually lacking. It is ironic that such an attitude may exist even among experimenters who fail to provide histological verification of the ablations they report in animals. It should be apparent that, if the aim of an experiment is localization of the lesion necessary and sufficient to produce a given symptom, then neither man nor animal in which locus is incompletely specified is a suitable experimental preparation. However, symptoms of brain damage raise a question other than localization, which is generally overlooked, namely, what is the nature of the alteration in behavior? The answer to this purely psychological question is relevant to theories of normal psychological organization. It

[1] This paper was read at the 1952 meeting of the American Psychological Association in a symposium entitled "Experimental and Clinical Psychology: The Application of Experimental Method to Clinical Material."

[2] Postdoctoral fellow, U. S. Public Health Service, 1952–53. The program of research described in this paper is supported in part by grants from the Commonwealth Fund of New York, and the Department of the Army, Office of the Surgeon General, under contract DA-49-007-MD-312.

is also an indispensable prerequisite to the adequate design and interpretation of localization studies. If the aim of an experiment is to describe and categorize precisely the nature of the behavioral change, then man is an eminently suitable experimental subject, and the fact that we cannot localize the causative lesion except grossly need not deter us from such an attempt. As Lashley (24) puts it, "The study of the effects of brain injury should serve as alternative to factor analysis, partialling out elementary functions or abilities which may vary independently."

What professional groups are in contact with material suitable for such studies, and how have they used it? Clinical psychologists are one such group, but so far have shown little interest in these problems. Among clinicians, evidence for the influence of the psyche on the soma is greeted enthusiastically (if not always critically), yet the complementary question of the influence of the soma on the psyche is neglected. The content of research in clinical psychology is typically the emotional "dynamisms" of the person; the usual aim is psychiatric diagnosis or therapy. Man as a rational animal is studied mainly by means of a few intelligence scales of an over-all character, the problem being generally conceived as how best to find a number which will represent the total intellectual attainment of the subject. There is little current research directed toward analysis of intelligence, and even less which makes use, in such analysis, of the leads offered by dissolution of complex nervous function.

Another group with opportunity to study the process of psychological disorganization consequent to damage of the nervous system is of course physicians, and some outstanding contributions have been made by a few neurologists, notably Hughlings Jackson, Head, and Goldstein. Their studies were not aimed at the discovery of diagnostic rules of thumb, but rather at the resolution of the factors underlying alterations of behavior or experience. All of them were greatly influenced by contemporary psychology, and it would be profitable to re-examine their notions in the light of present-day psychological theory. Their concepts require reformulation in operational terms, and their conclusions should be subjected to empirical test, using modern logical and statistical methods of inference.

Certain ideas and controversies, which have long been debated among neurologists, offer virgin territory for controlled experimentation. Jackson (17) held that the "organ of mind" is "only the most complex part of . . . a sensorimotor machine," that is to say, that even the highest centers represent parts of the body in terms of patterns of sensation and movement. This principle raises questions concerning the relation of sensation and movement to conceptual and symbolic processes, and to appropriate conduct. One of the more important of these questions to psychology is that of agnosia. Is it possible, for example, to lose visual recognition of objects in the absence of blindness or dementia?

Attempts have been made to attack the problem of agnosia in animals, with equivocal results. It is not clear, of course, that the syndrome could occur, or be recognized as such, in infrahuman species. Some first steps have been made, however, toward elucidating the nature of the relatively isolated defects in differential reactions which can be produced in animals by cerebral lesions. I should like to review briefly the outcome of these studies, and then discuss the reasons why further experimentation on man is likely to be necessary in order to resolve the issues.

According to traditional conceptions

of localization of function within the cortex, lesions of the "visuopsychic" band surrounding the visual sensory area should produce visual agnosia. Ablations of this region in monkeys have been studied independently by a number of investigators (1, 7, 8, 11, 24). Since this lesion involves risk of damage to the underlying optic radiations, unambiguous interpretation of any consequent abnormality demands histological demonstration of the intactness of the projection system. In the studies which have controlled this factor, no defects in visual function have been found, even though some of the problems required complex visual integrations, and were difficult enough to be near the limit of the monkey's capacity. Thus the ablation which on a priori grounds should be most likely to produce visual agnosia has failed to do so. I have attempted to produce tactile agnosia by restricted lesions of the "somatopsychic" area, and have likewise failed (5).

Although lesions of the areas supposedly concerned with sense-related ideas do not produce behavior resembling agnosia, ablations of cortex more remote from sensory areas may be followed by gross defects in discriminative behavior. Klüver and Bucy (19, 20) found that bilateral temporal lobectomy in monkeys produces undifferential approach responses to a wide variety of objects, regardless of whether these stimuli elicited approach, avoidance, or indifference in the normal animal. The syndrome also includes the tendency to examine all objects by mouth, acceptance of foods which are normally repellent, and greatly increased sexual activity.

Since this aggregation of abnormalities is difficult to derive from one which is fundamental, attempts have been made to dissociate the various defects by modification of the lesion. The temporal lobe includes two structurally distinct parts—paleocortex and neocortex. On the other hand, the temporal neocortex is structurally indistinguishable from that of the parietal and preoccipital regions (25). Chow, Pribram, and I (6) therefore considered that a reasonable preliminary approach to analysis of the syndrome might be to compare the effect of removing the parieto-temporo-preoccipital neocortex (excluding primary sensory areas) with the effects following traditional temporal or parietal lobectomies. The parieto-temporo-preoccipital operation was consistently followed by defects in visual discriminative learning, as measured by performances on formal habits of differential reaction to colors and patterns. All animals showed impairment on these tasks, although only half the animals showed behavior suggestive of the symptom which defines agnosia in man, that is to say, lack of customary reaction toward familiar objects.

Our rather extensive program of pre- and postoperative training and tests, and postmortem histological studies, allowed us to draw certain conclusions about the nature of the deficits on the formal visual discriminative tasks. In the first place, the deficits were something more than amnesias. The animals not only lost the habits to which they had been trained before operation, but had abnormal difficulty in reacquiring them. Second, the deficits were not caused by primary sensory impairment. Although some of the animals had primary visual defects consequent to invasion of the optic radiations, not all the animals suffered such damage. Impairment of visual discriminative habits was present in some animals in which the primary afferent system was completely intact, and in which there were no demonstrable scotomata or losses in acuity. Third, the deficits were specific to vision. Analogous somatosensory discriminative habits were much less

consistently and profoundly affected. Fourth, the deficits in visual discriminative learning were dissociable from gross spatial disorientation, from hypersexuality, and from aberrations of feeding behavior.

One of the questions raised by these results is why some of the animals showed impairment of object recognition and of orientation in space, and others did not. In our animals these abnormalities were present only in those individuals which had primary visual defects in addition to impairment of the ability required by the formal visual discriminative tasks. We know from the results of other studies, however, that scotomata and losses of acuity are in themselves insufficient to cause the symptoms of agnosia. Therefore, we hypothesized that two factors are necessary for the production of agnosia: One factor is related to damage to the higher levels of the sensory projection system, the other to lesion of "associative" cortex. This hypothesis recalls von Monakow's (26) view that permanent visual agnosia in man results only from combined lesion of both primary visual and other cortical areas. Additional data in support of such a two-factor theory might help to reconcile conflicting opinions concerning the basis of agnosia.

How to characterize either of these factors in the language of psychology, if they can be adequately described in such terms, is at present unknown, and is unlikely to be discovered from further animal experimentation. The first factor might be called "sensory," because it is the result of damage to the thalamocortical afferent system; but to what specific inabilities do we refer when we say "sensory defect"? We know that circumscribed damage to the striate cortex produces regional blindness, but we also know that this defect is not equivalent to that caused by peripheral damage (4, 23, 30).

The second factor is defined anatomically by lesion of "associative" cortex, and behaviorally by isolated impairment of visual discriminative learning. But how can we characterize in psychological terms a kind of deficit which is not visual, and yet is specific to vision? For example, normal monkeys learn color or brightness discrimination problems readily. A monkey with lesion of the temporal lobes, however, may be unable to learn to choose green in a red-green discrimination problem, although he immediately demonstrates a preference for red, and in that sense can discriminate. Monkeys with such lesions have abnormal difficulty in acquiring formal habits of discrimination between black and white, although they can easily distinguish a fine black thread against a black background.

In attempting to discover what the nature of this deficit may be, at least two major possibilities must be considered. The first is that what is lost is the capacity to use visual sense-data, which are themselves normal, as the basis of an acquired discriminative reaction; the second is that the sense-data themselves are altered in some subtle way.

Let us examine the first possibility that the deficit is an impairment of the capacity to use normal sense-data of a particular modality in acquired discriminative reactions. We might assume that each synaptic level in a sensory system represents a point of conjunction between the incoming sensory impulses and some class of reactions, and that the temporal neocortex is the point at which visual impulses influence most directly the processes on which the abilities required by formal discriminative tasks are based. These processes could still be activated by other modes of sensory stimulation, and consequently it could be shown, as in fact we did show, that discriminative learning in an-

other modality was not impaired similarly. With regard to various visual functions, the hypothesis predicts that whether they are lost or retained does not depend on the theoretical level of complexity of the function, but rather on the manner in which each is tested. It is the testing methods which are dissociable instead of the capacities supposedly tested. Much of the evidence from cases of agnosia in man points to the same conclusion. For example, recognition of an object is possible in one set of circumstances but not in another. Apraxia is similar in nature, as illustrated by Jackson's famous example of the patient who could not protrude his tongue on command, but who could lick crumbs from his lips (18). Both agnosia and apraxia can thus be conceived as a narrowing of the range of equivalence: in agnosia, restriction of response equivalence, and in apraxia, restriction of stimulus equivalence. I cannot emphasize too much, however, that there exists no adequate psychological formulation of why the responses, or the stimuli, are no longer equivalent.

The second possibility is that the isolated defects in discriminative reactions found in monkeys, and the symptoms of agnosia in man, are the result of a subtle alteration of the sense-data themselves. Head and Holmes (16) described the characteristic changes in tactile sensation produced by a cortical lesion as inconstancy of threshold, rapid local fatigue, hallucinations of touch, and abnormal persistence of sensations. Stein (29), working with cases of sensory involvement from central nervous lesion, showed that repetitive testing with von Frey hairs causes relatively great elevation of tactile thresholds. This may be true even of parts where the threshold appears normal to ordinary test. He suggests that difficulties in discrimination may be caused by this lability, that is to say, the first stimulus perceived may raise the threshold sufficiently so that a subsequent stimulus is not differentiable. Using Stein's methods, Cohen (9) examined several patients who showed defects of recognition with little or no sensory impairment, and found in all of them marked lability of thresholds and persistent after-sensations. The alterations of sensation which are revealed by the method of double simultaneous stimulation (2, 3, 10) may also be relevant to the problem of impairment of recognition of complex objects. If one stimulus "extinguishes" or "obscures" the perception of another, or displaces its subjective position, the resultant impression might be sufficiently different from normal perception to make recognition impossible. Assuming that such aberrations of the sensory process underlie the impairment of discriminative learning found in monkeys after lesions of "associative" cortex, a new conception of the function of these areas emerges. The traditional view is that impulses from the primary sensory areas are organized and patterned in "associative" cortex; instead, it may be that impulses from "associative" cortex play back upon the primary sensory areas, thereby modulating and stabilizing the sensory process.

It is apparent from this cursory survey of experiment and speculation concerning agnosia that the psychological problems presented by this syndrome are still unsolved. The animal experiments suggest the form of the solution, but they tell us little about the substance of the postulated factors. Even if we had the answer for the monkey, we still would not necessarily have it for man. A truly comparative attitude should prevent us from hasty generalization from one species to another even within the primate series. Whether the formulation developed in animal work is adequate for analysis of agnostic symptoms in man, and if so, how the

factors can be described in psychological terms, are empirical questions.

In collaboration with H. L. Teuber, S. Weinstein, and L. Ghent, I am attempting to develop empirical categories of somatosensory function in terms of symptom-complexes after brain lesions in man. To this end, men with proven loss of brain substance are given a series of tests of somatosensory and presumably related functions; all tests yield quantitative scores. The resulting profile is analyzed for relationships among symptoms, and compared to analogous profiles derived from the control groups, in order to clarify the nature of the alteration in the somesthetic process.

Population studied: The subjects of this study are drawn from the case load of the Psychophysiological Laboratory of New York University, gathered by H. L. Teuber and M. B. Bender for study of the after-effects of penetrating brain injury. These men are veterans, primarily of World War II, who have sustained combat injuries (gunshot wounds) of the brain or of peripheral nerves. They are ambulatory and report to the laboratory voluntarily for individual testing sessions. Most of these cases represent presumably stable lesions, in which further restitution of function, during the course of our survey, is unlikely to occur. The study of such stationary cases has obvious advantages for determination of the effects of loss of tissue per se. However, in order to extend the generality of our findings, we have recently begun testing also a group of cases in which benign tumors or encapsulated abscesses have been radically removed by the neurosurgeon.

Subgroups formed: The experimental group has brain injury and consequent somatosensory symptoms (as defined by the sensory tests in Group A, described below). There are three control groups, also made up of wounded veterans: one group with brain injury but without somatosensory symptoms; one group without brain injury but with somatosensory symptoms caused by peripheral nerve lesion; and one group without brain injury, and also without peripheral nerve lesion affecting the parts tested. Each group includes about thirty individuals.

Methods of obtaining data: The test program includes: (A) an evaluation by psychophysical methods of basic somatosensory functions, such as tactile sensitivity, cutaneous resolution of two points, and point localization. Measures of strength and coordination, and ratings of ability to perceive direction of passive movement of a joint, of tone, and of reflexes, supplement these data. (B) A quantitative study of aspects of the sensory process which are usually neglected, such as latency of sensation, adaptation time, duration of after-sensations, the effect on thresholds of prolonged testing, and of concurrent stimulation of other body parts. (C) Tests of discriminative ability, among forms, patterns, sizes, weights, textures, and grades of roughness. In these tests, the subject is given a series of models, and is asked to find duplicates from among an array. The procedure is carried out both unilaterally (the same hand feels the model and finds its duplicate) and bilaterally (one hand feels the model and the other hand finds the duplicate). (D) A comparison of the ability to solve problems and to derive concepts when the elements are perceived tactually, or visually. These tests include matching-from-sample, conditional matching, conditional reaction, abstraction, and sorting tests (12, 13, 14, 15, 21, 22, 27, 28). There is a visual form and a tactual form of each test. They will be described in detail in forthcoming publications; briefly, in each test the subject must discover a principle in order to respond correctly. The procedures are designed to minimize the effect of aphasic difficulties; they involve choice-reactions in a series of trials, approximating methods used in the animal laboratory. (E) Tests of orientation in external space and with respect to the body scheme. The subject is required to walk through complicated paths indicated on maps, perceived visually or tactually. Spatial sense, regarding the body, is tested by requiring the subject to touch in succession a series of points on the body

surface indicated on diagrams of front and back views of the body.

Methods of analyzing data: For each test, the performances of the groups are compared, and the differences evaluated statistically. The relationships among the various tests are determined separately for each group, by correlational techniques, in order to disclose patterns of normal and pathological function. Disparities among these group patterns will be examined for the purpose of arriving at an adequate description of the organization of somatic sensation at the cerebral level.

With this program, we hope to gain a clearer understanding of how somatic sensation and presumably related abilities are altered by central and peripheral lesion. We hope to find out if isolated disturbances such as those we produced in monkey occur in man, and if this is the case, what these disturbances mean. To what extent are certain somatosensory functions lost together, or preserved together, following cerebral lesion? Can discriminative function be disturbed independently of *any* alteration of the sensory process? Can there be disparate levels of ability to solve problems, dependent on the sense through which the relevant information is received and independent of sensory impairment? Can deficits in spatial "intelligence" be related to somatosensory or to visual defect? Can such deficits be specific to the body, or specific to external space? What systematic differences exist between normal patterns of sensory function and those found after lesions of the central, or peripheral nervous system? More specifically, are there differences in regional gradients of sensitivity, in the relation among various measures of sensory capacity, in temporal and spatial spread of effect of stimuli? The associations and dissociations of symptoms, which can be brought to light only by intensive study of the individual case, should contribute to our knowledge both of the basis of agnosia and of the normal gnostic process. The methods of classical experimental psychology and of the animal laboratory, when applied to the study of clinical material, may on the one hand yield more valid and reliable data than have hitherto been obtained with pathological cases, and on the other hand, richer and more understandable results than have been obtained with animals.

REFERENCES

1. ADES, H. W. Effect of extirpation of parastriate cortex on learned visual discrimination in monkeys. *J. Neuropath. exp. Neurol.*, 1946, 5, 60–65.
2. BENDER, M. B. Extinction and precipitation of cutaneous sensations. *Arch. Neurol. Psychiat.*, 1945, 54, 1–9.
3. BENDER, M. B., SHAPIRO, M. F., & SCHAPPELL, A. W. Extinction phenomena in hemiplegia. *Arch. Neurol. Psychiat.*, 1949, 62, 717–724.
4. BENDER, M. B., & TEUBER, H. L. Disturbances in visual perception following cerebral lesions. *J. Psychol.*, 1949, 28, 223–233.
5. BLUM, J. S. Cortical organization in somesthesis. *Comp. Psychol. Monogr.*, 1951, 20, 219–249.
6. BLUM, J. S., CHOW, K. L., & PRIBRAM, K. H. A behavioral analysis of the organization of the parieto-temporo-preoccipital cortex. *J. comp. Neurol.*, 1950, 93, 53–100.
7. CHOW, K. L. Effects of partial extirpations of the posterior association cortex on visually mediated behavior in monkeys. *Comp. Psychol. Monogr.*, 1951, 20, 187–217.
8. CHOW, K. L. Further studies on selective ablation of associative cortex in relation to visually mediated behavior. *J. comp. physiol. Psychol.*, 1952, 45, 109–118.
9. COHEN, G. Stereognostische Störung. *Dtsch. Z. Nervenheilk.*, 1926, 93, 228–244.
10. CRITCHLEY, M. Phenomenon of tactile inattention with special reference to parietal lesions. *Brain*, 1949, 72, 538–561.
11. EVARTS, E. V. Effect of ablation of prestriate cortex on auditory-visual association in monkey. *J. Neurophysiol.*, 1952, 15, 191–200.

12. GOLDSTEIN, K., & SHEERER, M. Abstract and concrete behavior. An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 7 (Whole No. 239).

13. GRANT, A. D., & BERG, E. A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J. exp. Psychol.*, 1948, **38**, 404–411.

14. HARLOW, H. F. Responses by rhesus monkeys to stimuli having multiple-sign values. In Q. McNemar and M. A. Merrill (Eds.), *Studies in personality*. New York: McGraw-Hill, 1942.

15. HARLOW, H. F. Solution by rhesus monkeys of a problem involving the Weigl principle using the matching-from-sample method. *J. comp. Psychol.*, 1943, **35**, 217–227.

16. HEAD, H., & HOLMES, G. Sensory disturbances from cerebral lesions. In H. Head, *Studies in neurology*. London: Hodder & Stoughton, 1920, Vol. 2.

17. JACKSON, J. H. Remarks on evolution and dissolution of the nervous system. In J. Taylor (Ed.), *Selected writings of. . . .* London: Hodder & Stoughton, 1932, Vol. 2.

18. JACKSON, J. H. Remarks on non-protrusion of the tongue in some cases of aphasia. In J. Taylor (Ed.), *Selected writings of. . . .* London: Hodder & Stoughton, 1932, Vol. 2.

19. KLÜVER, H., & BUCY, P. C. An analysis of certain effects of bilateral temporal lobectomy in the rhesus monkey, with special reference to "psychic blindness." *J. Psychol.*, 1938, **5**, 33–54.

20. KLÜVER, H., & BUCY, P. C. Preliminary analysis of functions of the temporal lobe in monkeys. *Arch. Neurol. Psychiat.*, 1939, **42**, 979–1000.

21. LASHLEY, K. S. The mechanism of vision. XV. Preliminary studies of the rat's capacity for detail vision. *J. gen. Psychol.*, 1938, **18**, 123–193.

22. LASHLEY, K. S. Conditional reactions in the rat. *J. Psychol.*, 1938, **6**, 311–324.

23. LASHLEY, K. S. Studies of cerebral function in learning. XII. Loss of the maze habit after occipital lesions in blind rats. *J. comp. Neurol.*, 1943, **79**, 431–462.

24. LASHLEY, K. S. The mechanism of vision: XVIII. Effects of destroying the visual "associative areas" of the monkey. *Genet. Psychol. Monogr.*, 1948, **37**, 107–166.

25. LASHLEY, K. S., & CLARK, G. The cytoarchitecture of the cerebral cortex of Ateles: a critical examination of architectonic studies. *J. comp. Neurol.*, 1946, **85**, 223–306.

26. MONAKOW, C. VON. *Die Lokalisation im Grosshirn*. Wiesbaden: Bergmann, 1914.

27. NISSEN, H. W., BLUM, J. S., & BLUM, R. A. Analysis of matching behavior in chimpanzee. *J. comp. physiol. Psychol.*, 1948, **41**, 62–74.

28. NISSEN, H. W., BLUM, J. S., & BLUM, R. A. Conditional matching behavior in chimpanzee; implications for the comparative study of intelligence. *J. comp. physiol. Psychol.*, 1949, **42**, 339–356.

29. STEIN, H. Die Labilität der Drucksinnschwelle bei Sensibilitätstörung. *Dtsch. Z. Nervenheilk.*, 1923, **80**, 57–74.

30. TEUBER, H. L., & BENDER, M. B. Critical flicker frequency in defective fields of vision. *Fed. Proc. Amer. Soc. exp. Biol.*, 1948, **7**, 123–124.

31. WEIGL, E. On the psychology of so-called processes of abstraction. *J. abnorm. soc. Psychol.*, 1941, **36**, 3–33.

[MS. received for early publication December 21, 1952]

Vol. 60, No. 3 May, 1953

THE PSYCHOLOGICAL REVIEW

PURPOSE AND LEARNING THEORY

OMAR K. MOORE

Tufts College
Systems Coordination Project, Naval Research Laboratory

AND DONALD J. LEWIS[1]

Northwestern University

We shall be concerned in this paper with the notion of "purpose," a notion that has long troubled social and biological scientists. It has been held by some psychologists (4) that a phenomenological observation of organismic behavior immediately reveals the essential goal directedness, the purposiveness, of this behavior, and that to talk of an animal's purpose in reaching a goal is legitimate as long as "purpose" is "defined" behaviorally. Other psychologists (2) have insisted vehemently that teleological concepts such as "purpose" either have no place in an objective account of behavior, or at most can be introduced only after they have been derived from primary principles. In this paper we shall attempt to show how such terms as "purpose" and "teleology" may be used with scientific respectability in connection with the primary principles of learning. In fact, we believe that these notions are essential to psychology, and, when properly used, make clear just what it is that the learning theorist is making laws about. It should be emphasized at the outset that no attempt is being made here to reintroduce into psychology mentalistic notions or entelechies that psychologists have labored so long to eliminate from their science.

We are aware that many persons, especially hard-headed experimentalists, wish to avoid any use of teleological concepts. This avoidance is understandable. Teleological concepts have been pre-empted in the past by those who have had little or no concern with the formulation of testable theory. Historically, "teleology" has been defined as the opposite of "mechanism." A teleological theory was taken to be one which explained the past in terms of the future. Purposes, ends, goals, or, in short, the terminal stage of any sequence of behavior acted in some unanalyzable way as one of the antecedent conditions necessary to reach the terminus. This usage of "teleology" has not proved useful in solving scientific problems.

It might be argued by some that if by "purpose" is not meant some metaphysical notion, the word should not be used at all, for it has too many mentalistic connotations. We feel justified in retaining the word, however, because this paper is concerned with goal behavior and with means and ends. But we are not going to talk about any animal as "having" a purpose any more than the engineers who work with

[1] The authors wish to thank Dr. David L. Olmsted and Dr. Richard Rudner for their careful reading of the manuscript and their helpful suggestions.

guided missiles that "seek" a target introject into the missile an entelechy or purpose.

We hope, in this paper, to redefine "purpose" and "teleology" in order to make them amenable to scientific psychological usage. We also hope to show the importance of these terms to learning theory and to show that the reinforcement theorists who have insisted most strongly that teleological concepts have no place in objective learning theory have made, paradoxically enough, essential use of the teleological frame of reference.

A significant methodological development in recent years has been a series of reformulations of the concept "teleology." Certain philosophers of science, principally Singer (3), Churchman and Ackoff (1), and also a number of scientists, prominent among whom are Wiener (6), and Rosenblueth and Bigelow (5), have been active in this effort. Their work has been extremely helpful in enabling us to make an analysis of the relationship between learning theory and teleological concepts. Our point of view is closest to that of Churchman and Ackoff. In the short space available it will be impossible to give a full account of their views or to show in just what way their views differ from ours.

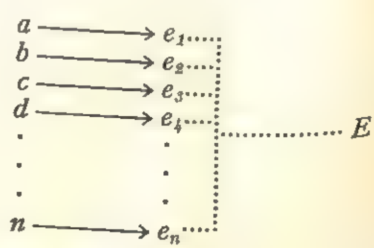
Two Frames of Reference

There are at least two quite different, yet compatible, frames of reference that may be profitably employed in the analysis of behavior. They are the mechanical and the teleological (but not in the historical sense previously noted). When a scientist uses the mechanical frame of reference, ideally at least, he attempts to specify the conditions, both necessary and sufficient, to bring about some state of affairs. For purposes of illustration, let us use the following expression:

$$a \longrightarrow e_1$$

If the set of conditions designated by "a" is sufficient to bring about the state of affairs designated by "e_1," then whenever a occurs, e_1 occurs. If a is a necessary condition for e_1, then e_1 will not occur unless a has occurred. As a matter of fact, scientists have seldom, if ever, been able to specify the sufficient conditions for any occurrence. Nonetheless, from the point of view of the mechanical frame of reference, the ideal to be achieved is the specification of the necessary and sufficient conditions for any event.

In the following diagram, let us assume that a is a necessary and sufficient condition for e_1, b for e_2, c for e_3, and d for e_4. This means that when a occurs, e_1 will always occur, and if a does not occur, e_1 will not occur. An explanation using the mechanical frame of reference may be said to have been given when a set of well-confirmed laws has been formulated which will enable one to predict the occurrence or nonoccurrence of an e_1 given the occurrence or nonoccurrence of an a.



There is another feature of the mechanical frame of reference that requires comment. It is essential to specify just what the antecedent and consequent conditions are in any sequence of events. Two ways of specifying objects and environments that have been distinguished especially concern us here. One is the physical classification and the other is the morpho-

logical. According to Churchman and Ackoff,

> In physical classification we group objects or events (behavior patterns) in terms of a quantified property expressed along a physical scale, such as temperature, weight, wave length, velocity, etc. Morphological classification, on the other hand, is a non-physical method of classifying. It may take either a quantitative or qualitative form. When quantitative, the scientist chooses a range along a physical scale as a criterion for membership in the class. . . . We may also have qualitative morphologies, as when we specify "red objects," or "heavy objects," or "rapid walking" (1, p. 35).

The difference between physical and morphological classification is not that between precision and vagueness, but is the difference between the specification of a unique value along a scale and the specification of a range along a scale. The limits of any given range may be very precisely stated.

The ideal use of the mechanical frame of reference requires at least the physical classification of antecedent and consequent conditions. Otherwise it would be extremely difficult to specify the necessary and sufficient conditions for some occurrence. In our illustration, let us for the moment assume that a does not occur and b does. Under these conditions e_2 would appear. But if e_1 and e_2 are not distinguished on the basis of a physical classification, then e_1 might be confused with e_2 and a would seemingly fail as a necessary condition for e_1.

Means-ends. The teleological frame of reference, as developed in this paper, is concerned with the notion of means-ends. Therefore the problem to which we now turn our attention is that of explicating this notion. We want to be able to answer the following question: Under what conditions may some specified occurrence be considered as a means to some end? In terms of our illustration, we shall want to be able to answer the question as to when a, b, c, or d may be treated as means to an end. Let us begin by considering e_1, e_2, e_3, and e_4 as falling within the same morphologically defined class, the range of which is at least extensive enough to include these four elements. Let us assume further that if at least one—and it makes no difference which one or ones—of these possible occurrences obtains, then E obtains. "E" is the name of the state of affairs that obtains when at least one of the four events (e_1, e_2, e_3, e_4) occurs. More precisely, "E" is a variable which takes names of states as instances of substitution. We require a name for such states of affairs in order to avoid awkward circumlocution. The question now arises as to the relationship between a, for instance, and E. It can be seen that a is not a necessary condition for E because even though a does not occur, E may occur through the occurrence of b, c, or d. However, a is the sufficient condition for E. That is to say, if a occurs, e_1 occurs and thus E obtains. What this analysis of the teleological frame of reference provides is an explication of the notion of means, and alternative means, to some specified end. The relationship of a to E is that of a being a means to E. Of course b, c, and d are also means to E, that is, means to the same end. The four alternatives can be described as potential means.

In comparing the two frames of reference, there are several important differences to be noted. First, in using the teleological frame of reference, it is permissible, although not necessary, to make a morphological classification of consequent conditions. It is not essential to be able to differentiate between, for instance, e_1 and e_2 since no matter which one occurs, E obtains. The successful employment of the me-

chanical frame of reference requires that both the antecedent and consequent conditions be classified at least physically. Second, in the mechanical frame of reference, the antecedent state of affairs is both a necessary and sufficient condition for the consequent state. In the teleological frame of reference, any given antecedent state is not a necessary condition for the achievement of the consequent condition, namely, E in our example, although it is a sufficient condition.

Functional class. At this point it is necessary to make one further distinction. In order to do this we introduce the notion of a functional class. Since a, b, c, or d have been treated as means or potential means to E, they can be thought of collectively as members of the same class. The criterion for membership in this class is that of being a potential means to some specified end. Thus, a, b, c, or d may differ greatly physically and/or morphologically and yet have the property of being means to some specified end, and thus collectively constitute a functional class.

Reformulations of "Teleology"

Our object, in setting forth some of the essential features of the teleological frame of reference, is to enable us to make clear in just what way learning theory of the reinforcement variety *does* implicitly make use of this frame of reference. Reinforcement theorists seek to avoid any use of teleological concepts. In fairness to their position, it should be pointed out that they are opponents of that doctrine of teleology which leads directly to vitalism and emergentism, and which involves the introjection of an entelechy or purpose into the organism itself. Hull (2, p. 27) has suggested a way to avoid such subjective concepts. "This is to regard, from time to time, the behaving organism as a completely self-maintaining robot, constructed of materials as unlike ourselves as may be." He believes that the tendency on the part of the observer to impute an entelechy, soul, spirit, or demon into such a robot is less likely than if the scientist were observing a living organism. He also states that the robot concept is an effective prophylaxis against the tendency to reify a behavior function. "To reify a function is to give it a name and presently to consider that the name represents a thing, and finally to believe that the thing so named somehow *explains* the performance of the function" (2, p. 28).

We, of course, are no more anxious than Hull to reify functions or to introject entelechies into men or robots. It would seem, however, that the scientists who have been most intimately connected with the theory and construction of servomechanisms and sequence-controlled calculators have found it profitable to make extensive use of teleological concepts. Perhaps it would be more accurate to say that they have divested these concepts of their demoniacal connotations and rendered them usable for scientific purposes. Wiener (6) would characterize Hull's self-maintaining robot as active, purposive, teleological, and as capable of high-order predictions. One might assume that Wiener has simply repeated the ancient fallacy of introjecting purpose into the machine itself. Quite to the contrary, his analysis—behavioristic in character—concerns only the input and output of the machine. He says, "Given any object, relatively abstracted from its surroundings for study, the behavioristic approach consists in the examination of the output of the object and the relations of the output to input" (6, p. 1). To say that a machine is purposive in Wiener's sense is not to say that it has some mysterious entelechy within it. The actions of the

machine may be assessed with respect to their being means toward some end. In so far as these actions are thus describable they may be said to be purposive.

The cybernetic definitions of teleological concepts have been framed with reference to the problems of interpreting the behavior of certain classes of machines and neural systems. This has resulted in a conceptual model which is workable for the problems of cybernetics. However, the data of current learning theory requires certain modifications of these concepts. Churchman and Ackoff (1) have recently framed a more widely applicable set of definitions which we shall draw upon for our present purposes. Wiener develops the idea that an object behaves purposively only if it pursues the same goal by modifying its behavior under varying conditions. The actions of a thermostat, for example, may be viewed as purposive in that the thermostat turns on the furnace when the room temperature drops below a critical point, and when the temperature goes above this critical point, it turns off the furnace. The end or goal in this example is the maintaining of the temperature at a critical point, and the actions of the thermostat may be viewed as a means of achieving that end in a changing environment. Let us, however, take for an example a naïve hungry rat placed in a Skinner box. This animal may perform a variety of responses and may, after a time, obtain food. But because the conditions in the box have remained constant, we cannot consider the rat's behavior purposive under the Wiener definition. Thus we are in the uncomfortable position of having to maintain that behavior of the thermostat is purposive, but the behavior of the rat is not.

To avoid this difficulty, we have worked out (following Churchman and Ackoff [1]) certain teleological categories: (*a*) Extensive Function, (*b*) Intensive Function, and (*c*) Purposive Function. When using the teleological frame of reference one may classify any behavior as belonging to one of these categories. (*a*) If X (any object) accomplishes some specified end by displaying relatively invariant behavior in a wide range of environments, then X may be said to have an extensive function. (*b*) If X accomplishes some specified end by changing its behavior if and only if the environment changes, but exhibits only one type of behavior in any given environment, then X may be said to have an intensive function. (*c*) If X accomplishes some specified end by exhibiting different types of behavior, whether the environment changes or remains the same, then the actions of X may be said to be purposive. According to this schema, the behavior of the thermostat would be classified as having an intensive function; the behavior of the rat in the problem box as purposive.

Hullian Theory and "Purpose"

Let us turn our attention now more particularly to the Hullian formulation of learning theory. We wish to show that this theory makes essential use of the concept of purpose as just explained. In order to make our case, we must be very clear about *essential use*. To be precise about this extremely important point we must make the distinction between a set of lawlike statements and what they are about. Any theoretician has at least a twofold task: (*a*) He must classify the objects, events, or situations which are to be the subject matter of his laws, and this implies that he have a principle of classification. (*b*) He must formulate a set of lawlike statements, which have as their subject matter the objects, events, or situations that he has classified, and these

statements should systematically relate parts of the classified subject matter to other parts. Ideally, these lawlike statements will enable him to predict the occurrence of the classified happenings, and thus will acquire the status of laws. The learning theorist can properly be said to make essential use of teleological concepts if these concepts are employed either in his classificatory system or in his laws.

Hullian theorists have devoted their attention almost exclusively to the formulation of testable laws. But we have already noted that the theoretician has a twofold task and that he must do more than formulate laws. He must also classify the objects, events, or situations which will be the subject matter of his laws, and he must make clear the principle of this classification. It is this second task that the Hullians have neglected. Apparently they are unaware of what principle of classification they have used, and it is partly because they have neglected making clear this principle that the controversy about "teleology" exists.

Our point is that the Hullians have been using a classificatory principle, and that it is a teleological one of the purposive variety. Now that we have made this assertion, the question arises as to how it can be proved or disproved; what sort of evidence would be required to confirm or disconfirm it? Quite obviously we should turn to the actual work of the learning theorist. Let us take Hull's book *Principles of Behavior* (2) for illustration. We find that Hull, on the one hand, presents a number of experiments, and, on the other, he formulates a set of laws to explain what happens in these experiments. The question is: On what basis does he select these experiments as subject matter for his laws? We submit that each and every one of the experiments cited in his book involves a situation in which the behavior of the organism is assessed with respect to its efficacy in accomplishing certain ends.

We find, for example, that Demonstration Experiment A on page 70 is concerned with a rat placed on a grid which is subsequently electrified. Hull points out that the rat displays a variety of responses, one of which, the leaping over a barrier, removes it from the electrified grid. Hull classifies the responses of the rat into two categories, the futile and nonfutile. What can Hull possibly mean by characterizing a response as futile? It is not a physical or morphological property of a response to be futile or nonfutile, and the notion of futility or nonfutility is certainly not part of the mechanical frame of reference. We believe that what he means, or should mean, is that certain of the responses of the rat, when placed on an electrified grid, are means to a specified end, and other responses are not. The end which Hull selects is need reduction. The rat's leaping over a barrier is a means to the end, need reduction. There is a further employment of the teleological frame of reference by Hull in this experiment. In a series of trials, the rat makes a number of escapes from the charged grid. Hull pays no attention to the exact location from which the rat leaps or to the specific musculature involved (for the rat may leap from varying stances) when he describes the rat's escape. If he were making a purely mechanical analysis, he would, of course, have to do this very precisely. Nor does he pay attention to the exact spot where the rat lands on the other side of the barrier, and this is something else that would have to be done precisely in a purely mechanical analysis. Let us assume that the rat may leap from any one of four physically different loca-

tions, and land on any of four physically different spots across the barrier.

$$\left.\begin{array}{l} \text{Location } a_1 \longrightarrow \text{Spot } e_1 \\ \text{Location } a_2 \longrightarrow \text{Spot } e_2 \\ \text{Location } a_3 \longrightarrow \text{Spot } e_3 \\ \text{Location } a_4 \longrightarrow \text{Spot } e_4 \end{array}\right] \cdots\cdots E$$

Hull is perfectly willing to count a landing at any one of these four spots as constituting an escape from the grid. What this means, then, is that he is interested in the relationship between a_1, a_2, a_3, a_4, which can be classified morphologically, and E. These four starting locations, and the appropriate jumping reactions, constitute members of the same functional class, any one of which is a means to the same end, namely, escape from the grid. The behavior of the rat on the electrified grid may be categorized as purposive. The rat accomplishes some specified end by exhibiting different types of behavior whether the environment changes or remains the same.

In "Demonstration Experiment C" (2, p. 75) Hull cites a conditioned-reflex learning experiment in which a dog learns to lift its foot in response to a buzzer, with shock serving as the unconditioned stimulus. Hull says, "No doubt the dog makes many other muscular contractions in addition to those which result in the lifting of the foot, but these are usually neglected in such experiments." Why, we might ask, are they neglected? It is quite evident that learning theorists are not merely interested in a purely mechanical analysis of what happens when an organism is subjected to certain stimuli. Rather, they wish to construct a set of laws that will enable them to predict the occurrence of those responses that are relevant to the achievement of a specified goal.

In the short space available we cannot cite all the experiments performed by reinforcement theorists, or even all of those mentioned in Hull's book. We are confident, however, that an examination of the experiments performed by reinforcement theorists will reveal the same properties. What these experiments have in common is that they deal with objects that can exhibit different types of behavior whether the environment changes or remains the same. And the experimenter singles out of the myriad of responses that might be investigated just those that are relevant to the achievement of some specified end. What the experiments have in common, then, is the use of the teleological frame of reference, and even within that frame of reference, the use of the special category of purpose. For example, the experimenters do not treat behavior in terms of the categories of either extensive function or intensive function. The experiments are not performed on objects that display relatively invariant behavior in a wide range of environments, as do clocks, for example. Nor do they involve organisms or machines that exhibit different behaviors in different environments, but only one type of behavior in any given environment, as do amoeba and some servomechanisms. It should be clear at this point, then, that the learning theorists are making systematic use, as any theorist must, of some principle of classification in selecting subject matter about which to formulate laws. The Hullian learning theorists have never made their principle of classification clear.

TOLMANIAN THEORY AND "PURPOSE"

It may seem that the position that has been set forth in this paper is equivalent to the one Tolman and his followers have long urged. Certainly Tolman has talked about the importance of purpose, ends-means relationships, goals, and has even said that "behavior as behavior reeks of purpose

and of cognition" (4, p. 12). However, Tolman's use of teleological concepts is quite different from ours. To the Tolmanian, purpose is something that the animal has within it, and that an independent, neutral observer ascribes to the organism as a result of an analysis of its behavior. According to this view, we can speak, upon observing a cat in a problem box, of the "cat's purpose of getting to the outside, by bursting through the confinement of the box," or of the "cat's purpose as a determinant of the cat's behavior" (4, p. 13). To Tolman "The rat accepts or rejects and persists to or from food, blind-alleys, true path sections, electric grills, etc., only in so far as these latter function for him as subordinate goals . . ." (4, p. 28). In other words, goals are goals *to* the rat, and purposes are purposes *of* the rat. In our analysis, *purpose* is a function of the classificatory system employed by the investigating scientist; it is never some faculty inside the organism.

The Tolmanian analysis differs from our own in another important respect. In his system the distinction between the classificatory schema employed to select the subject matter for the laws and the laws of the system themselves is obscured. It seems to us that Tolman uses the word "purpose" both as the name of a phenomenon to be explained and as an explanatory device. This in itself is a procedure of dubious scientific value. Ultimately it is a viciously circular procedure. Moreover, it makes it extremely unlikely that a scientist can successfully carry out the twofold task that we have delineated above.

The learning theory developed by the reinforcement theorists, in our opinion, is not guilty of this confusion. It is, however, beyond the scope of the present paper to examine the relationship between the Hullian laws and the implicit Hullian classificatory system in detail.

In summary, many critics have charged that reinforcement theory is defective because it does not take purposive behavior into consideration, and proponents of reinforcement theory have admitted the absence of teleological elements and have taken this absence as an index of the scientific character of their theory. Following the lead of certain cyberneticians and philosophers of science, we have worked out a new formulation of the teleological frame of reference which avoids metaphysical entanglement, and have shown that the reinforcement theorists have implicitly been making use of this frame of reference. The Hullian laws are set up to explain the conditions under which an organism learns. From the implicit Hullian point of view, to say that an organism has learned something is to assert that there is a high probability that the organism will exhibit responses that are a means to the end need reduction. Thus we have seen that the concept of learning itself is one that can be meaningfully used only within the teleological frame of reference.

REFERENCES

1. CHURCHMAN, C. W., & ACKOFF, R. L. Purposive behavior and cybernetics. *Soc. Forces,* 1950, 29, 32–39.
2. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
3. SINGER, E. A., JR. *Mind as behavior.* Columbus: R. G. Adams, 1924.
4. TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.
5. ROSENBLUETH, A., WIENER, N., & BIGELOW, J. Behavior, purpose, and teleology. *Phil. Sci.,* 11, 1943, 18–24.
6. WIENER, N. *Cybernetics.* New York: Wiley, 1948.

[MS. received April 14, 1952]

THE PROBLEM OF "WHAT IS LEARNED?"

JAN SMEDSLUND

Institute for Social Research, Oslo

In a recent paper (2) Kendler discusses the problem of "what is learned?" and concludes that it is a pseudoproblem stemming from the error of reifying theoretical constructs. It seems to us that Kendler's interpretation of the formulation, "What is learned?" as being synonymous with the formulation, "Which is the best theoretical model for representing the learning process?" is not a very reasonable one. We think there is another more important and operationally meaningful interpretation.

We interpret the formulation "what has the organism (O) learned in the series of situations (S)?" as being synonymous with the formulation "what changes have occurred in O's behavior and learning in any situation, whether it belongs to S or not, as a function of O's being presented to S?"

If the situations belonging to S are (approximately) identical, as has usually been the case in learning experiments, then one cannot infer anything about what has been learned, because (*a*) in a constant situation there is no possibility of knowing which cues or aspects of the proximal stimulation have been reacted to and, what is more important, (*b*) one cannot know on which distal variables the learning has been focused. The situation is a special case of what Brunswik (1) has called "channeled mediation" and "artificial tying" of variables.

This type of experimental design has often resulted in a failure to recognize the distal focusing of behavior, and has given rise to the frequent, but never very consistent, efforts to maintain a proximal-peripheral classification of variables, involving the belief that what is learned are movements in response to proximal stimulation.

As the learning situations become more variable, one may infer more about what has been learned in these situations. As the series of situations becomes more representative of the organism's ecological universe, it becomes easier to infer what changes have occurred in the ecological universe and, consequently, to infer on which distal variables the learned behavior is focused.

The proximal-peripheral classification of variables and the resulting primitive concept of what is learned, i.e., something that may be directly inferred from a constant situation, have led to the pseudoconcept of transfer. The procedure of determining what is learned in a series of situations (S) and the procedure of determining transfer from S are identical. They consist of varying the situation and of recording what changes in behavior and new learning have taken place as a function of the learning in S. Therefore the concept of transfer becomes unnecessary. The problem of predicting transfer is the problem of predicting what will be learned.

In addition to being superfluous, the concept of transfer can be shown to be an artifact created by the traditional experimental design. If transfer is defined as effects of learning in one concrete situation or group of situations on behavior and on learning in any other concrete situation or group of situations, then every instance of learning outside the psychological laboratory is also an instance of transfer, because every observation of learning requires at least two situations and every two situations, except in extremely well-controlled experiments, have at least some differences. Thus, when there is no artificial tying of variables, the concepts of learn-

ing and of transfer cannot be distinguished.

Let us now consider the problem of predicting *if, how fast,* and *what* a given organism or group of organisms will learn in a given series of situations. In our opinion the relevant variables are all central, and because these variables are the result of earlier learning processes, we arrive at the conclusion that, in general, every prediction of learning must, explicitly or implicitly, be based on a diagnosis of relevant parts of what the organism has learned before. In the limiting case of the newly born, the determinants of the learning process are not the result of earlier learning processes, but are nevertheless central.

By the statement that all variables relevant to the learning process are central we mean that they are variables of (earlier formed) systems that have to be inferred from a complex set of observations and that they are consequently not external. There is considerable evidence in literature (partly discussed in [4]) indicating that no constant relationships between simple external measures and the learning process can be found, except within extremely narrow fields. The motivational and perceptual determinants of the learning process in a given external situation are a function of earlier learning processes and the influence of the time factor (distribution of practice, delay of reinforcement, etc.) likewise changes with learning. We believe our point of view to be related to that of Young (5), who likewise points out the inadequacy of physical (i.e., simple externally anchored) measures.

An example of how the learning process is clearly a function of dimensions of central structures would be the case of a group of Communists and a group of anti-Communists memorizing the content of a speech given by Stalin. Here the relevant central structure is the attitude toward Communism. Another example may be taken from animal learning. Studies of "place learning" and of "reasoning" (Maier) indicate that the rat, at the age when it is employed in learning experiments, has already formed what Piaget (3) has called "spatial groupings." This means that, given the information that a rat possesses the kind of central structure called "spatial grouping" and with such and such values of the parameters of this structure, one may predict (given certain additional information) if, how fast, and what this rat will learn in a given spatial context. Indices of the existence of a "grouping" of spatial relations are the ability of coordinating separate experiences, making detours, reversing paths, inferring directions, etc.

We conclude that the problem of what is learned, as defined above, has a clear-cut operational meaning and that it is an important one in the psychology of learning, because any successful theory of learning presupposes (*a*) that methods of diagnosing central structures (what has been learned earlier) have been developed and (*b*) that the laws relating the learning process to the dimensions of these earlier formed structures have been found.

REFERENCES

1. BRUNSWIK, E. *Systematic and representative design of psychological experiments.* Berkeley and Los Angeles: Univer. Calif. Press, 1949.
2. KENDLER, H. H. "What is learned?"—A theoretical blind alley. *Psychol. Rev.*, 1952, **59**, 269–277.
3. PIAGET, J. *La construction du réel chez l'enfant.* Neuchatel et Paris: Delachaux & Niestlé, 1937.
4. SMEDSLUND, J. A critical evaluation of the current status of learning theory. *Nordisk Psykol. Monografiserie*, 1952, 2.
5. YOUNG, P. T. The role of hedonic processes in the organization of behavior. *Psychol. Rev.*, 1952, **59**, 249–262.

[Received for early publication January 15, 1953]

Psychological Review
Vol. 60, No. 3, 1953

A SINGLE THEORY FOR REMINISCENCE, ACT REGRESSION, AND OTHER PHENOMENA[1]

ELI SALTZ

Technical Training Research Laboratory,
Human Resources Research Center[2]

The theoretical formulation here presented was at first designed simply as an attempt to explain in terms of Hullian constructs the act regression phenomenon reported by Hamilton and Krechevsky (9), Kleemeier (21), O'Kelly (32) and others. The specific theoretical model used, in fact, is a modification of Spence's transposition theory. It soon became evident, however, that the model might become useful not only for the prediction of regression phenomena and transposition, but also for certain features of retroactive inhibition, spontaneous recovery, and disinhibition. If upheld by further research, the theoretical model gives signs of being a very useful tool in the unifying of several areas of behavior study.

It should be emphasized that the theoretical model here presented attempts to deal only with a limited segment of behavior. The model is not to be considered an attempt to predict all of behavior.

Theoretical Model

The application of the stimulus generalization principle to the two processes of excitation and inhibition as previously used in discrimination learning (39) provides a mechanism in terms of which predictions can be made in the areas of instrumental learning, verbal learning, and conditioning.

Eight assumptions will be used to indicate the nature of this mechanism. The specific functional relations required to state the assumptions precisely are not crucial in our present lack of knowledge of the interactions exhibited in the empirical results.

The eight assumptions involved in the model are:

1. Each time a particular response R_1 to a given stimulus S_0 is rewarded, a hypothetical excitatory tendency of S_0 to elicit R_1 is increased. The amount of this excitatory tendency connecting S_0 to R_1 will henceforth be referred to as the *Absolute Excitatory Strength* of S_0 to elicit R_1.

This assumption is widely held by psychologists. McGeoch (28), in his chapter on the law of effect, summarizes the verbal learning data on this point and finds strong evidence in favor of a reinforcement position. Hull (15) summarizes the data on this topic in instrumental learning and conditioning, and he comes to the same conclusion in this area.

2. Each time a response R_1 is elicited, either overtly or covertly, by a stimulus S_0 and no reward occurs, a hypothetical inhibitory tendency of S_0 not to elicit R_1 is increased. The amount of this inhibitory tendency preventing the occurrence of R_1 upon the appearance of S_0 will henceforth be referred to as the *Absolute Inhibitory Strength* of S_0 to R_1.

The experimental work relevant to this assumption has been limited in

[1] The writer wishes to express his gratitude to Professor H. P. Bechtoldt for his encouragement during the formulation of the theoretical model presented in this paper, and to Professor G. Robert Grice who read and criticized the manuscript for the present paper.

[2] The opinions or conclusions contained in this paper are those of the author. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

amount. Melton (30, 31) suggests that in verbal learning some factor other than the competition between responses seems to be operating in those situations in which a second response (R_2) is learned in place of the original response (R_1); he calls this additional factor "Factor X" and considers it an unlearning factor due to the occurrence and nonreward of R_1 to the stimulus when R_2 is the correct response. Melton points out that if competition between the two responses is the only consideration leading to the "forgetting" of R_1, then the effect of R_1 competing with R_2 should be as great as the effect of R_2 competing with R_1; therefore the decrement to R_2 due to proactive inhibition should be as great as the decrement to R_1 due to retroactive inhibition. Since the retroactive-inhibition decrement was much greater than the proactive-inhibition decrement, Melton concluded that both an unlearning factor and a competition factor contribute to the decrement of R_1 upon the learning of R_2. The fact that few overt intrusions of R_1 occur during the learning of R_2 can be made consistent with a sizable decrement in R_1 due to unlearning, Melton suggests, by assuming that R_1 responses occur covertly during the learning of R_2.

Operationally, extinction in both instrumental learning and in conditioning can be defined as a function of nonreinforcement; consequently, it appears safe to state that nonreinforcement produces a weakening of the S-R connection in both these areas.

3. The absolute excitatory strength of S_0 to elicit R_1 and the absolute inhibitory strength of S_0 upon R_1 summate algebraically. The algebraic difference between the absolute excitatory and inhibitory strengths of an S_0-R_1 connection will henceforth be referred to as the *E-I Strength* of S_0 to elicit R_1.

4. When two incompatible responses, R_1 and R_2, compete for elicitation, the difference in *E-I* strengths (determined algebraically) of the two tendencies is effective in determining the response made. The strength of R_1 as a function of the difference between the *E-I* strength of R_1 and R_2 will henceforth be designated the *Effective Strength* of R_1.

Evidence for assumption 4 is suggested by the results reported by Melton and Irwin (30), McGeoch (26), and Siipola and Israel (37). All of these investigators suggest that increasing the number of presentations of R_2 decreases the strength of R_1, though the results of Melton and McGeoch seem to suggest that this relationship is not necessarily monotonic when extreme overlearning of R_2 occurs.

Assumptions 3 and 4 both involve statements of linear relationships between excitatory and inhibitory strength. Probably the relationships are more complex than simple linear functions. However, linearity seemed the best possible guess at this stage of knowledge. Spence (38) reports the assumption of algebraic summation of excitatory and inhibitory strength to be satisfactory as a first approximation.

5. Stimuli other than S_0 acquire a tendency to elicit and to inhibit the response R_1 as a result of the generalization of the hypothetical absolute inhibitory and absolute excitatory tendencies along several stimulus similarity dimensions. The curves will be termed the *Generalization Gradients* for inhibition and excitation.

Evidence supporting the concept of generalization of absolute excitatory strength in verbal learning has been provided by Yum (44), Dulsky (3) and others.

Lashley and Wade (24) have recently attacked the existence of a generalization gradient of excitatory strength in instrumental learning. However, replies by Hull (16), Grice

(4, 5), Grice and Saltz (6), and others to the Lashley and Wade paper appear to make the assumption of such a gradient a reasonable one.

Generalization of absolute inhibitory strength has not been directly investigated with verbal learning materials; when nonverbal stimuli are used, however, there seems to be little doubt as to the occurrence of the phenomenon. Hull (16) summarizes these data.

The following three assumptions are not implicitly or explicitly stated in Spence's transposition theory.

6. The slope of the curve for generalization of inhibition is steeper than the slope of the curve of excitation.

Assumption 6 concerning the relative slopes of the gradients of generalized absolute excitation and inhibition has few experimental data to back it. Hovland (13) reports that in conditioning there appears to be a tendency for inhibition to drop off more quickly than excitation as degree of generalization is increased.[3] However, there appear to be no data relevant to this assumption in the area of verbal learning.

7. After the removal of a stimulus, its trace persists for a brief period; the trace grows progressively weaker with the passage of time.

The above is obviously a restatement of the stimulus-trace hypothesis of Pavlov (33) and of Hull (15). The trace hypothesis was rephrased at this time to de-emphasize the neurological aspects of the formulations found in both Hull and Pavlov.

The eighth assumption is merely a statement of a specific condition of assumption 5 concerning generalization of excitation.

8. Passage of time, after the cessation of an activity, is accompanied by an alteration of "set" stimuli within the subjects. The original set stimuli can be recovered by a warm-up period consisting of activity resembling the original activity.

The evidence for the above assumption arises largely from the work of Irion (17, 18). Irion found that memory loss between learning and relearning could be decreased if, just prior to relearning, colors were presented to the subjects in a fashion similar to the manner in which the verbal material was originally presented for learning. Irion interpreted these results as indicating that the delay between original learning and relearning was accompanied by a loss of "sets" acquired during original learning. This constitutes a stimulus alteration.

The fullest description of the more basic interactions between the above variables is presented in the following section on act regression. The discussion of the relationship of other experimental phenomena to the model will simply allude to the discussion presented in the act regression section.

INSTRUMENTAL LEARNING

1. *Act Regression*

In the classical act-regression studies, one response (e.g., turning left in a T maze) is established to a particular stimulus and then is extinguished while a competing response (e.g., turning right in a T maze) is developed to the same stimulus. During the acquisition of the second response a new stimulus (e.g., a momentary electric shock) is presented at the choice point before the response has been made in that particular trial. Upon presentation of the new stimulus the animals often tend to exhibit the first (extinguished) response

[3] It should be noted that while Hovland's verbal conceptualization of inhibition is different from that proposed in this paper, the two inhibition concepts are isomorphic in terms of the operational definition offered in this paper and the operational definition implicit in the Hovland experiment.

more frequently than the more recently learned second response.

Hamilton and Krechevsky (9), Sanders (36), Martin (29), O'Kelly (32), and Kleemeier (21), among others, have used fundamentally the above experimental design. Both Martin and Kleemeier have shown that regression tendency increases with a greater relative number of reinforcements to the first response. O'Kelly has demonstrated that reducing the animal's drive level (satiating the animal) has the same effect as electric shock in producing regression.

In terms of the theoretical model, an absolute excitatory tendency, R_1 (turning left), is developed to S_0 (the choice-point stimulus) by means of repeated reinforcements. Next R_1 is no longer

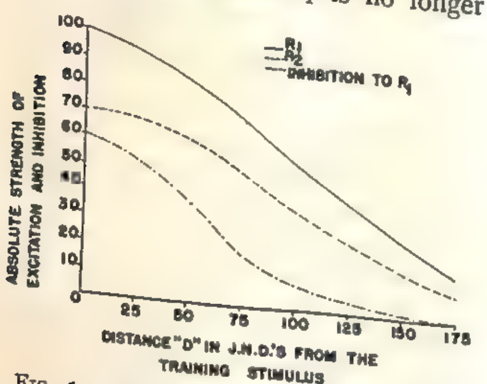


Fig. 1. Hypothetical generalization gradients of excitation and inhibition.

For illustrative purposes, the equation for generalized excitation was taken as $E_j = E_e^{-.000025d^2}$ and the equation for generalized inhibition was taken as $I_j = I_e^{-.00008d^2}$. E_j and I_j represent excitatory and inhibitory strengths, respectively, at a generalized stimulus j; d is the distance in *j.n.d.*'s between j and the training stimulus; and e is a constant equal to 10. These parameters are arbitrarily selected and are not intended even as approximations of the correct parameters. However, any parameters should show the same tendencies for interaction characteristics between responses as the above parameters, as long as both excitation and inhibition generalization gradients fall monotonically, and as long as the inhibition gradient falls more rapidly than the excitation gradient.

reinforced upon presentation of S_0 and some incompatible response, R_2 (turning right), is established instead. Repeated nonreinforcement of overt and covert R_1 responses results in the development of inhibition to R_1 at the training stimulus S_0, while repeated reinforcements of R_2 lead to the growth of absolute excitatory strength to R_2. At the point where the *E-I* strength of R_1 falls below that of R_2, S_0 will elicit R_2 rather than R_1. The occurrence of any modification of S_0 at this point will change the stimulus situation for the subjects so that some stimulus S_j, different from S_0, will be present. Figures 1 and 2 illustrate the effect of this state of affairs upon the effective strength of R_1.

As the stimulus changes in j.n.d.'s we find that the *E-I* strength of R_1 to the generalized stimuli falls less quickly than does the *E-I* strength of R_2. This differential reduction in the *E-I* strengths follows from the consideration of the generalized strengths of excitation and inhibition associated with R_1 and the strength of the excitation tendency to R_2. Since the slope of the inhibition gradient is such that inhibition falls more rapidly than excitation, the loss of absolute inhibition to R_1 is greater than the loss of absolute excitation as the two move along the stimulus alteration dimension; therefore, the *E-I* strength of R_1 will, up to some point, increase with stimulus alteration. As the inhibition to R_2 is zero, the *E-I* strength of R_2 will drop consistently as a function of stimulus alteration.

Figure 2 shows the effective strength of R_1 at the various stimulus generalization points and shows how increased stimulus changes, at least up to some point, tend to allow for greater and greater effective strength of R_1; this same tendency occurs whether the absolute strength of R_1 is greater or less than that of R_2, but the stronger R_1 is

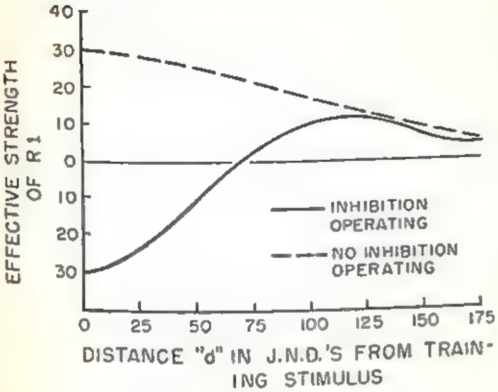


FIG. 2. Hypothetical generalization gradients of excitation affected by inhibition and not affected by inhibition when inhibition generalizes less than excitation. The values used to plot the above figure were obtained from Fig. 1.

in absolute strength in relation to R_2 the greater will be the effective strength of R_1. Thus the two findings that have emerged from the regression studies are both predictable on the basis of the present theoretical model: (*a*) Shock at the choice point, altered hunger state, and altered set conditions lead to regression. (*b*) The greater the number of reinforcements of R_1 over R_2 during the original learning the greater the amount of regression.

2. *Transposition*

A classical example of transposition is one in which a subject first learns to respond to the larger of two stimuli; when the smaller of the two stimuli is removed and a new stimulus which is larger than the original large stimulus is introduced, the subject may respond to the larger of the new pair just as he responded to the larger of the old pair. In those cases where the subject reacts to the "relationship" between sizes, going to the larger stimulus despite the fact that this was not the stimulus originally rewarded, transposition is said to have occurred. Such "relationship" responses to new stimulus pairs, as can be readily seen, are possible whenever the stimuli can be arranged along a dimension. Brightness has also been used quite often to demonstrate this effect.

The assumptions, in the theoretical model presented in this paper, which are not part of Spence's (39) transposition theory do not appear relevant to transposition. Consequently, the predictions originally made by Spence also follow from the present theory.

Spence hypothesizes that the reward member of the pair of stimuli generalizes excitation; the nonreward member generalizes inhibition. He then predicts from the interactions of these two processes that transposition is greatest to pairs of stimuli close, on the stimulus dimension, to the originally learned pair. Pairs progressively farther from the original stimuli should show the effect less strongly. Finally, pairs at the extreme of the dimension should show an inversion of the relationship response. Several experimental studies tend to sustain both the predictions of the inversion of the relationship response (8, 41, 43) and of the progressive decrease in tendency toward transposition as a function of distance from the original training stimuli (7, 22, 39).

However, the results of Kendler (20) are not completely consistent with the last stated prediction. Kendler measured the percentage of transposition along a brightness dimension. She trained one set of animals to respond to the brighter of two stimuli, then presented them with test pairs which were brighter than the training pair. The percentage of transposition first decreased in accordance with the prediction; however, without reaching a stage in which the percentage of transposition was that to be expected by chance, the curve began to rise again. These results are contrary to Spence's original prediction. However, they can be ex-

plained in part at least by the findings of Grice and Saltz (6); Grice and Saltz tested Hull's stimulus dynamism postulate and found that increased intensity of the stimulus resulted in increased response tendency. Kendler's discrepant results occurred when she presented stimuli brighter than the training pair. The relationship response was the one to the brighter stimulus. This stimulus dynamism postulate suggests that increasing brightness raises the response tendency. Thus the interpretation can be made that the tendency toward the relationship response was increased at the extremes of brightness used by Kendler.

Kuenne (23) has shown that the transposition follows the Spence predictions in preverbal children, but that in older children the relationship response appears to be stable and independent of the distance between the test pair of stimuli and the training stimuli. She interprets this as indicating that words take on specific cue value and that older children do not respond to the visual stimuli as such but rather that they respond to verbal cues like "bigger than" or "smaller than."

3. *Spontaneous Alternation*

Dennis (2) reports that when animals are given two alternative paths of equal length, both paths ending in reward, the animals will show a greater-than-chance tendency to avoid the path chosen on the preceding trial. The longer the paths are made, the stronger the alternation tendency becomes. This is a reasonable result, in terms of the theoretical model presented above. Delay of reward is a point along the reward-no-reward continuum. In fact, any response is eventually followed by some reward, though this reward may not arise until hours after the response has been made; in such a situation the reward may probably be considered as not effective for reinforcing the particular response. Thus, a choice point followed by a long path to reward should result in inhibition accruing to the path chosen; if the two alternative paths are similar in appearance at the choice point, both the rewarding and inhibiting properties of the chosen path will generalize to the previously ignored path. Since inhibition generalizes less than reward, the previously ignored path will have a greater *E-I* strength than the previously selected path.

Passage of time has been assumed to alter the animal's set, and this results in generalization of both excitation and inhibition. Since inhibition generalizes less than excitation, it is predicted that the alternation tendency should decrease as a function of the time interval between the two trials. This is exactly what Heathers (10) found.

The more the two alleys were dissimilar in appearance at the choice point, the less generalization of either excitation or inhibition there should be between them. Consequently, the alternation tendency should be reduced. If the choice-point response is followed by immediate reward, no inhibition should develop and consequently alternation should occur no more frequently than would be expected by chance. The greater the amount of reward, the less alternation should occur as the *E-I* strength of the first alley will be increased. These three predictions, as far as the author knows, have never been tested. They would contribute considerable evidence concerning the theoretical model.

VERBAL LEARNING

1. *Retroactive Inhibition*

As can be seen by examining the paradigm of act regression discussed above, the retroactive-inhibition paradigm and the act-regression paradigm are ex-

tremely similar. The act-regression technique can be described operationally as a modified retroactive-transfer paradigm (A-B, A-K, A′-B′) in which the modification in the "relearning stage" involves a change in the stimulus conditions.

As a consequence of the theoretical model presented above, it is predicted that when, in the relearning phase of a retroactive-inhibition situation the stimulus member is altered within certain limits, the relearning should be faster than when the identical stimulus is presented.

2. *Strength of Retroactive vs. Proactive Inhibition as a Function of the Interval Between the Learning of A-K and the Relearning Phase*

Melton and Von Lackum (31) found that after learning A-B followed by A-K it was more difficult to relearn A-B than to relearn A-K. This, it will be recalled, was used by Melton and Von Lackum as proof for the existence of an "Inhibition" factor.

Underwood (42) repeated the Melton and Von Lackum experimental design using paired associates and using two different time intervals between the A-K and the relearning phase. Basically, one group relearned after 5 hours, another group after 48 hours delay. His results were the same as those of Melton and Von Lackum in the 5-hour delay group: The A-B list was more difficult to relearn than the A-K list. However, this was no longer true after 48 hours delay. The 48-hour delay group showed no significant difference in relearning for A-K over A-B.

The Underwood data become predictable in terms of the theoretical model if Irion's warm-up effect experiments are brought into the picture. If the interval between learning and relearning is accomplished by a stimulus change, the amount of change being a function of the length of delay, then Underwood's 48-hour group would relearn A-K and A-B under a more altered stimulus complex than would his 5-hour group. As a consequence, it is predicted from the theoretical model that A-B would be stronger after 48 than after 5 hours. This, fundamentally, is what Underwood found.

3. *Bowed Serial-Learning Curve*

The typical curve of item difficulty in a serial-learning list of homogeneous items is bowed in appearance. An item somewhat closer to the end of the list than to the beginning is the most difficult to learn. Difficulty falls off in a monotonic fashion in both directions from the most difficult item. Consider A-B-C-D-E-F-G as constituting a serial list, each letter representing a separate item of the list. The correct response to any item of the list is the item immediately to its right. However, in addition to the association A-B, the remote forward associations A-C, A-D, etc. are formed, as has been shown by McGeoch (27), among others. And in addition to incorrect forward associations, McGeoch (27) and others have shown that backward associations develop; item D, for example, has a tendency to evoke items C, B, and A as responses. These remote associations may be considered as products of the stimulus trace. The trace of an item occurs simultaneously with some remote list item. When the remote item is reinforced to its correct stimulus, this reinforces the connection between the trace and the remote item. In the present paper, both correct responses after the response word appears and correct anticipations will be considered reinforcing states of affairs. The overt and covert occurrence of incorrect forward and backward associations is not rewarded, and increments of inhibition develop to prevent these responses from occurring to incorrect stimuli.

This inhibition generalizes along a similarity dimension so that the strength of the response is decreased not only to incorrect stimuli but also to correct stimuli.

If certain assumptions are made concerning the relative strengths of forward and backward remote associations, the frequency of occurrence of items as incorrect responses to other list items, when plotted against serial position of the responses, will produce a curve identical in shape to the typical serial-learning curve. To the extent that the above analysis is pertinent to the prediction of the bowed serial-learning curve, the items in the most difficult-to-learn serial position should be associated more strongly to incorrect stimuli than any other item. The item in the most easily learned serial position should be associated least often to incorrect stimuli. The present writer tested these predictions in a study to be reported in more detail at a later date, and the predictions were sustained.

4. *Reminiscence in Serial Learning*

The passage of time produces an altered stimulus condition, with inhibition generalizing less than excitation. Therefore, following a rest after learning much of the inhibition will no longer be effective. As the inhibition is greatest in the central portion of the serial list, delay after learning will produce an increment in response strength of these center items. The loss of inhibition will produce an increased tendency toward false responses at the ends of the serial list; thus, more correct responses should occur at the center of the list, fewer at the ends of the list than before the rest period. These predictions correspond to the results obtained by Hovland (14).

Irion's (17) discussion of the warm-up effects indicates that, several trials after the end of the rest period, the *S*s' original set conditions are recovered. Consequently, the prediction can be made that after a warm-up of several trials the center of the serial-learning curve would again drop in number of correct responses; this should occur because the warm-up should return the *S*s to their original set stimulus conditions where the inhibition effects are active. The Hovland data bear out this prediction also (14).

The prediction can be made that reminiscence effects will be due to recovery of words from the center of the serial list. Whether the *total* number of words recovered will be greater during the reminiscence trial than during the practice trial depends on the interaction between inhibition dissipation with passage of time and the interference effects caused by this dissipation. Generalization due to passage of time will dissipate the inhibition which develops in items due to specific non-reinforcements; it also dissipates the inhibition which accrues to items due to generalization of inhibition from similar items in the list. Since the exact gradients are unknown, it is difficult to make predictions about the optimal conditions in which total number of words during the reminiscence trials will be greater than the total number during the last practice trial. An interesting variation on the reminiscence studies has been attempted. This variation consists of presenting a meaningful passage for subjects to read. After the passage has been read, a series of questions on the passage is presented. Some of the questions quote the passage, others paraphrase it. Equating the two types of questions for difficulty immediately after reading it, Checov (1) finds that, after one day, the paraphrased questions are answered correctly more frequently than the direct quotation questions. From the standpoint of the theoretical model, the para-

phrased material is an alteration of the original material and produces the generalization necessary for reminiscence.

CONDITIONING

1. *Disinhibition*

Pavlov (33) reports that an extinguished conditioned response often reappears when its conditioned stimulus is accompanied by an extraneous stimulus. This he calls disinhibition. In terms of the theoretical model, the extinction of a response is the nonrewarding of the response until the absolute inhibitory strength is approximately equal to the absolute excitatory strength. An extraneous stimulus alters the stimulus complex. As inhibition does not generalize as much as does excitation, the conditioned response will tend to reappear.

The prediction from the model is that as the extraneous stimulus alters the total complex more and more, the conditioned response should wax, then wane in strength. The waning results from an alteration of the complex to a point along the stimulus dimension where the absolute excitatory strength of the response is extremely small. These predictions coincide with the results reported in Pavlov (33).[4]

A more direct test of the model is possible at this point. If, instead of introducing an extraneous stimulus after extinction of the conditioned response, the conditioned stimulus is altered, disinhibition should result.

2. *Spontaneous Recovery*

Pavlov (33) reports that an extinguished conditioned response reappears on the day after extinction. The reappearance of the response, as predictable from the model, is a result of the altered set stimulus which occurs as a function of time. Since inhibition does not generalize as much as does excitation, the alteration of the set stimulus produces a generalization situation in which the excitatory strength of the response is greater than the inhibitory strength.

Also arising from the model is the prediction that learning a new response, since this involves a stimulus change, should facilitate spontaneous recovery. This, in essence, is the result reported by Liberman (25).

3. *Inhibition of Delay*

When a long period consistently elapses between the onset of the conditioned response and the onset of the unconditioned stimulus, the conditioned response moves in time of appearance from immediately after the conditioned stimulus to immediately before the unconditioned stimulus.

In terms of the model, since there was a delay of reinforcement to the response when the response occurred immediately after the conditioned stimulus,[5] inhibition accrued to the response. However, when the response occurred after the conditioned stimulus plus a time delay, it was followed more quickly by reward (the unconditioned stimulus) and was fixated here.

The theoretical explanation provided above for spontaneous recovery is applicable here. On a subsequent day after training of the delayed response, the appearance of the conditioned stimulus should result in an immediate appearance of the conditioned response. Rodnick (34) found this to be the case.

[4] The similarity of this explanation of disinhibition to those of Pavlov (33) and of Hull (15) should be noted.

[5] The initial occurrence of the conditioned response immediately after the conditioned stimulus, in such delayed-reinforcement situations, is partly an artifact of training. The delayed conditioning is difficult to establish. Consequently, a shorter interval between conditioned and unconditioned stimuli is usually used at the beginning of training, and this interval is increased as training progresses.

The prerequisite variables for disinhibition as explained above are also present in this formulation of inhibition of delay. As a consequence, the prediction can be made that an external stimulus, accompanying the conditioned stimulus, should result in an immediate appearance of the conditioned response. Rodnick reports this phenomenon (35).

4. *Massing of Practice*

In dealing with the effect of massing on conditioning, the discussion in this paper will be modeled on the preceding discussion of massing in verbal learning. The essentials of that thesis involve a response becoming attached to "incorrect" stimuli because traces of the incorrect stimuli are present while the response is being rewarded; the incorrect S-R connections are inhibited and this inhibition generalizes to the correct S-R connections; massing produces stronger traces of incorrect stimuli since the trace strength is a function of time after removal of the stimulus.

In a similar manner, massing in conditioning may be thought of as producing connections between the conditioned response and incorrect stimuli. The conditioned response *itself* can be considered as producing a trace stimulus which can be attached to the next occurrence of a conditioned response; lack of reinforcement (or delay of reinforcement) produces inhibition to this S-R connection. In terms of this explanation, the inhibition originally develops between the response trace (which acts as a stimulus) and the conditioned response; however, the inhibition moves back in the conditioning sequence and eventually occurs at the onset of the conditioned stimulus. This is because the response trace is a conditioned response occurring at the time of onset of the conditioned stimulus, and as such it becomes a mediating stimulus for the inhibition.

As a consequence of the above discussion, it is predicted that, under massed conditions, multiple conditioned responses (e.g., two or more consecutive conditioned eye blinks following the conditioned stimulus) will increase in frequency as a function of number of trials until increments of inhibition with further trials extinguish them.

One question arises immediately concerning the trace-response connection discussed above: What reinforcing state of affairs occurs which allows such a connection to form? To be consistent with the reinforcement position taken earlier in this paper, it is necessary to indicate the existence of some reinforcement for the connection between response trace and succeeding response. The answer to this question is fairly obvious. In Hullian terms, the conditioned stimulus is a primary reinforcing agent. Since this follows the trace-response occurrence, it reinforces this occurrence.

Two factors are, as implicit in the discussion so far, involved in the intertrial interval: reinforcement and inhibition. The longer the interval (i.e., the less the massing of practice) the more the inhibition to the trace-response connection. This is true because longer intervals mean longer delay of reinforcement since the S_c-S_u of the next trial constitutes reinforcement. For the same reason, the longer the interval, the less the reinforcement of the interfering trace-response connection. Extreme massing, then, should reduce the inhibition to the interfering response and might be expected actually to facilitate conditioning. Fundamentally, this is what Hilgard and Marquis (11) report that Calvin found. Nine trials per minute result in poorer learning than do three per minute. However,

18 trials per minute result in better learning than do nine per minute.[6]

Also predictable from the model are the results reported by Jones (19), who found that at the beginning of practice after an interval, conditioning performance is better than during the original massed training. Such time intervals, it will be recalled, have previously in this paper been assumed to produce a set stimulus change; stimulus change permits inhibited responses to recur.

5. *Switzer Effect*

Switzer (40) reports that during the first few extinction trials following massed conditioning, an *increase* in rate of conditioned-response evocation occurs. This result is easily explainable in terms of the regression model. Extinction trials constitute a modification in the stimulus patterns and cause a generalization of the excitatory strength of the S-R connection. They also cause generalization of the inhibition arising from the trace-response connection being nonrewarded. Since inhibition generalizes less than excitation, an increment in *E-I* is to be expected.

Hovland (12) has examined the Switzer effect in several different sets of experimental conditions. Knowing the results obtained by Hovland it is possible to make specific assumptions about the shape of the curves of excitation and inhibition so that Hovland's results could be predicted. However, the regression model in the form stated above is not applicable to an explanation of Hovland's data. While it might be interesting to make specific statements concerning the variable parameters, the writer feels that this is outside the scope of the present paper.

[6] These results are of particular interest in that the prediction was made before the writer was aware of Calvin's experiment.

CONCLUDING REMARKS

The present paper has outlined a modification of Spence's transposition theory in an eight-assumption model. The model was then shown to be relatively consonant with past research in the areas of instrumental learning, verbal learning, and classical conditioning. The writer is aware that such a demonstration can show only the possibility that a theory may have merit. Only the actual use of the model for designing new experiments can tell us how much predictive value the model has. Several such predictions were attempted in the body of this paper.

The writer is also aware that, despite his Hullian bias, the model does not fit neatly into Hullian theory. For example, Hullians have used the concept of work inhibition to explain such phenomena as spontaneous recovery, reminiscence, and spontaneous alternation. The present writer has dealt with these phenomena without using the concept of work inhibition. The extent to which either position is "correct" in this conflict (and indeed the extent to which there is a conflict) must await further research before a decision can be made. In general, a serious scientist must recognize the unfortunate probability that no theory proposed in psychology for some time to come will be exempt from at least some modification. A slow, bit-by-bit alteration of previous ideas appears to be the way in which scientific models of respectable predictive power develop.

REFERENCES

1. CHECOV, L. An investigation of reminiscence as a function of type of learning material and level of difficulty. Doctoral dissertation in psychology, Univer. of Washington, 1951.
2. DENNIS, W. Spontaneous alternation in rats as an indicator of the persistence of stimulus effect. *J. comp. Psychol.*, 1939, 28, 305–312.

3. DULSKY, S. G. The effect of a change of background on recall and relearning. *J. exp. Psychol.*, 1935, **18**, 725–740.
4. GRICE, G. R. The acquisition of a visual discrimination habit following response to a single stimulus. *J. exp. Psychol.*, 1948, **38**, 633–642.
5. GRICE, G. R. The acquisition of a visual discrimination habit following extinction of response to one stimulus. *J. comp. physiol. Psychol.*, 1951, **44**, 149–153.
6. GRICE, G. R., & SALTZ, E. The generalization of an instrumental response to stimuli varying in the size dimension. *J. exp. Psychol.*, 1950, **40**, 702–708.
7. GULLIKSEN, H. Studies of transfer of responses: I. Relative versus absolute factors in the discrimination of size by the white rat. *J. genet. Psychol.*, 1932, **40**, 37–51.
8. GUNDLACH, R. H., & HARRINGTON, G. B. The problem of relative and absolute transfer of discrimination. *J. comp. Psychol.*, 1933, **16**, 199–206.
9. HAMILTON, J. A., & KRECHEVSKY, I. Studies in the effect of shock upon behavior plasticity in the rat. *J. comp. Psychol.*, 1933, **16**, 237–253.
10. HEATHERS, G. L. The avoidance of repetition of a maze reaction. *J. Psychol.*, 1940, **10**, 359–380.
11. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
12. HOVLAND, C. I. "Inhibition of reinforcement" and phenomena of experimental extinction. *Proc. nat. Acad. Sci.*, 1936, **22**, 430–433.
13. HOVLAND, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.
14. HOVLAND, C. I. Experimental studies in rote-learning theory: VI. Comparison of retention following learning to same criterion by massed and distributed practice. *J. exp. Psychol.*, 1940, **26**, 568–587.
15. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
16. HULL, C. L. The problem of primary stimulus generalization. *Psychol. Rev.*, 1947, **54**, 120–134.
17. IRION, A. L. Retention and warm-up effects in paired associate learning. *J. exp. Psychol.*, 1949, **39**, 669–675.
18. IRION, A. L., & WHAM, D. S. Recovery from retention loss as a function of amount of pre-recall warm-up. *J. exp. Psychol.*, 1951, **41**, 242–246.
19. JONES, H. E. The retention of conditioned emotional responses in infancy. *J. genet. Psychol.*, 1930, **37**, 485–498.
20. KENDLER, TRACY S. An experimental investigation of transposition as a function of the difference between training and test stimuli. *J. exp. Psychol.*, 1950, **40**, 552–562.
21. KLEEMEIER, R. W. Fixation and regression in the rat. *Psychol. Monogr.*, 1942, **54**, No. 4 (Whole No. 246).
22. KLÜVER, H. *Behavior mechanisms in monkeys.* Chicago: Univer. of Chicago Press, 1933.
23. KUENNE, MARGARET R. Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, **36**, 471–490.
24. LASHLEY, K. S., & WADE, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1939, **25**, 261–272.
25. LIBERMAN, A. M. The effect of interpolated activity on spontaneous recovery from experimental extinction. *J. exp. Psychol.*, 1944, **34**, 282–301.
26. MCGEOCH, J. A. The influence of degree of interpolated learning upon retroactive inhibition. *Amer. J. Psychol.*, 1932, **44**, 695–708.
27. MCGEOCH, J. A. The direction and extent of intra-serial associations at recall. *Amer. J. Psychol.*, 1936, **48**, 221–245.
28. MCGEOCH, J. A. *The psychology of human learning.* New York: Longmans, Green, and Co., 1942.
29. MARTIN, R. F. "Native" traits and regression in the rat. *J. comp. Psychol.*, 1940, **30**, 1–16.
30. MELTON, A. W., & IRWIN, J. M. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.
31. MELTON, A. W., & VON LACKUM, W. J. Retroactive and proactive inhibition in retention: Evidence for a two-factor theory of retroactive inhibition. *Amer. J. Psychol.*, 1941, **54**, 157–173.
32. O'KELLY, L. I. An experimental study of regression. II. Some motivational determinants of regression and perseveration. *J. comp. Psychol.*, 1940, **30**, 55–95.

33. PAVLOV, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep.) London: Oxford Univer. Press, 1927.

34. RODNICK, E. H. Characteristics of delay and trace conditioned responses. *J. exp. Psychol.*, 1937, **20**, 409–425.

35. RODNICK, E. H. Does the interval of delay of conditioned responses possess inhibitory properties? *J. exp. Psychol.*, 1937, **20**, 507–527.

36. SANDERS, M. J. An experimental demonstration of regression in the rat. *J. exp. Psychol.*, 1937, **21**, 493–510.

37. SHIPOLA, ELSA M., & ISRAEL, H. E. Habit interference as dependent upon stage of training. *Amer. J. Psychol.*, 1933, **45**, 205–227.

38. SPENCE, K. W. Analysis of formation of visual discrimination habits in chimpanzee. *J. comp. Psychol.*, 1937, **23**, 77–100.

39. SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

40. SWITZER, S. A. Backward conditioning of the lid reflex. *J. exp. Psychol.*, 1930, **13**, 76–97.

41. TAYLOR, H. A study of configuration learning. *J. comp. Psychol.*, 1932, **13**, 19–26.

42. UNDERWOOD, B. J. Retroactive and proactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, **38**, 29–38.

43. WARDEN, C. J., & WINSLOW, C. N. The discrimination of absolute versus relative size in the ring dove *turtur risorius*. *J. genet. Psychol.*, 1931, **39**, 328–341.

44. YUM, K. S. An experimental test of the law of assimilation. *J. exp. Psychol.*, 1931, **14**, 68–82.

[MS. received April 18, 1952]

Psychological Review
Vol. 60, No. 3, 1953

OBSERVATIONAL DEFINITIONS OF EMOTION

WILSON McTEER

Wayne University

In the theoretical discussion of the topic of perception, psychologists for generations have emphasized that the perceived object is a function of the observer's "set to perceive." Strangely, this same relationship is often overlooked when investigators come into disagreement in defining perceptual objects in other areas of psychological investigation. While it is generally realized that in any research one investigator can observe only a portion of the many potentially observable sequences which are occurring, it is too frequently assumed that because one observer reports on an event of a named category, then another investigator under similar objective circumstances can and will observe the same sequence of events (if his interest is directed toward that same category). If the second investigator does not verify the observations of his predecessor, all too frequently charges and countercharges are exchanged with regard to (*a*) the objective circumstances (experimental situation) or with regard to (*b*) the fallibility of the observer; rarely is there an attempt made to cross-check the (*c*) observational set of the two investigators.

In general, such clashes occur with least frequency in those areas in which precise description of experimental situation and apparatus narrows the range of set variability to almost zero, as in the study of sensation or of the conditioned reflex. On the other hand, such disagreements occur with regularity in those areas which are concerned with the study of human behavior in molar situations, that is, in those areas in which set variability permits a wide range of differential selection of significant events out of the composite observed (emotion, motivation, personality, social adjustment, as examples).

The writer has been interested for some years in the controversies concerning the concept of emotion. In this area in particular, "set to perceive" is clearly responsible for much of the diversity of factual and theoretical observations. Even though P. T. Young in his *Emotion in Man and Animal* (24) gave one of the most inclusive definitions of this term which has yet been written, he creates confusion in that he strives to include in a single observational description three fundamentally different perceptual viewpoints. Likewise, D. O. Hebb in his article "Emotion in Man and Animal: An Analysis of the Intuitive Processes of Recognition" (11), gives a striking example of the manner in which different perceptual sets produce data of unlike utility in studying emotional behavior in the chimpanzee. In this article he reported that in the Yale Laboratory of Primate Biology in Florida a

> . . . thoroughgoing attempt to avoid anthropomorphic description in the study of temperament was made over a two-year period. . . . A formal experiment was set up to provide records of the actual behavior of the adult chimpanzees, and from these records to get an objective statement of the differences from animal to animal. All that resulted was an almost endless series of specific acts in which no order or meaning could be found. On the other hand, by the use of frankly anthropomorphic concepts of emotion and attitude one could quickly and easily describe the peculiarities of the individual animals (11, p. 88).

This latter type of description, he asserted further, provided "an intelligible and practical guide to behavior" of the individual animals which was not found in the description of the separate acts.

Considerably earlier, in the contrasting summaries presented by Landis (16) and Bard (1) in Murchison's *The Foundations of Experimental Psychology* (1929), it was evident that differing perceptual sets were producing remarkably unlike data (yet not necessarily conflicting, and all quite properly classified under the same general topic of emotion). Later, this disparity of data led to the series of articles by Dashiell (6), Dunlap (10), Meyer (19), and Duffy (8, 9) with the common theme that the term *emotion* no longer had any proper use within the frame of scientific psychology.

Nevertheless, it is the contention of this paper that much of the seeming confusion and controversy concerning the concept of emotion would be removed if the differing perceptual sets of the contributors were recognized.

If the logic of the following schematic analysis is adopted, there are only three fundamentally different points of view which are available for use in the observation of emotional phenomena. However, the simplicity of this analysis is reduced when it is recognized that within the frame of each of these points of view there are possible significantly different perceptual sets and, within each of these sets, multiple levels of description (microscopic to macroscopic) may further complicate the observational report.

These three fundamental points of

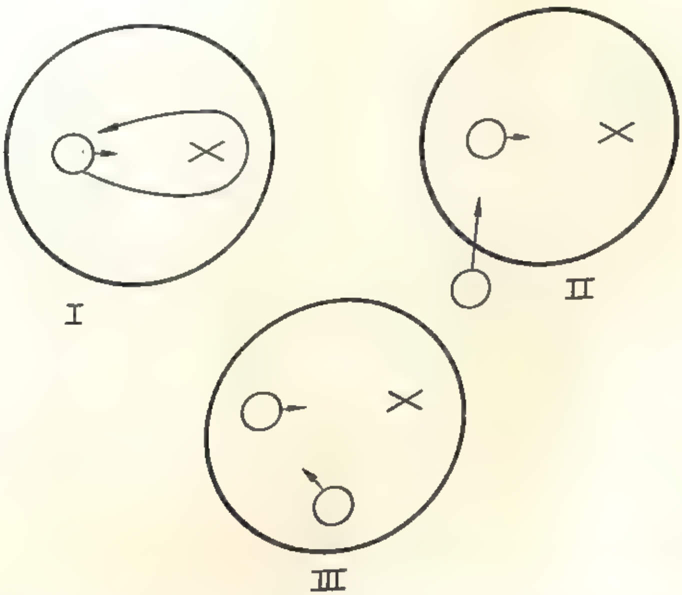


FIG. 1. Schema illustrating observational viewpoints in the study of emotion. In the above diagrams the large circle is intended to represent the extent of the effective stimulating environment impinging upon an emotional organism at any moment of observational study (obviously the cumulative effect of the stimulations of previous moments is not shown). The organism which is being observed is indicated by the small circle within the larger one. The dominant aspect of the immediate environment which is the focus of the organism's emotional attack or withdrawal is represented by an X. A second small circle with an arrow attached indicates both the reporting observer and the direction of his observation.

view are presented in the accompanying diagrams. They may be identified as: I. Introspection; II. Objective Observation; and III. Participant Observation. They are to be distinguished one from the other in terms of the relationship of the observer to that which is being observed.

INTROSPECTIVE OBSERVATION

In Diagram I is presented the point of view of the introspective observer. Here, the emotional organism and the observer are the same individual, and, as a result, this position is subject to all of the limitations that arise from any attempt to carry on two highly conscious activities at the same time. This introspective approach was the method used by the earlier writers who were criticized by William James (14) in his famous Chapter on Emotion and it is likewise the point of view which was used by William James himself. James differed from the earlier writers only in that his perceptual set was directed toward the observation of organic and sensory changes rather than toward the vaguer feelings and mental reactions of his predecessors. Lange (15) differed from James in that, with his medical background, he was set to observe changes of the capillary blood vessels as being the significant phenomena. Dewey's (7) contemporary alternate hypothesis indicated a perceptual set to observe emotion in relation to ongoing personal activities rather than in relation to internal physiological changes.

OBJECTIVE OBSERVATION

In Diagram II is presented the point of view of the objective observer who keeps himself outside of the stimulating environment affecting the emotional organism. The Gesell observation dome, or the more recent technique using the one-way visual screen to hide the observer, typifies close-range objective observation of this nature. Telescopic viewing, examination of motion picture or other graphic time records, or the reviewing of protocol or diary accounts, permits a similar objectivity at a greater distance (either spatial or temporal).

Historically, it may be indicated that this objective position was the unwitting point of view taken by Sherrington (22) and Cannon (2) in their early research inspired by the James-Lange theory of emotion. Where James had appealed to introspective observation, Sherrington and Cannon resorted to operative experiments with animals. Consequently, their observations on emotion were of necessity objective, even though the frame of original reference had been introspective.

Regrettably, even this objective point of view permits significant variations in perceptual set in observing emotional sequences. Many investigators have focused chiefly upon the physiological changes evoked by experimental environments which were presumed to be uniform (that is, uniformity of the external situation was stressed, while little or no consideration was given to the varied life history backgrounds which the *Ss* brought with them to the experimental situation). Cannon's (3) later experiments typify this physiological perceptual set. Other investigators have focused more upon the uniformities and variations of facial and gestural expression when members of the same species were exposed to supposedly uniform emotional stimulations, as did Landis (16) in his classical study of human facial expressions. Yet others, perhaps not identified as psychologists, have investigated the temporal antecedents or causes in the life environment which have led to similar emotional consequences in certain life situations. Injury-accident prevention studies in factories, traffic accident rec-

ords and analyses in our city and state traffic bureaus, and the sickness-death probability studies by insurance companies all fall within the pattern of objectivity proposed here; and all have been directed toward the reduction of potentially emotional events in the life experience of many citizens.

An interesting outgrowth of an emphasis upon extreme objectivity in observing emotional phenomena was the denial by the behaviorist Weiss (23) of the significant existence of emotional phenomena, since the concepts of feeling and emotion had their reference origin in introspection.

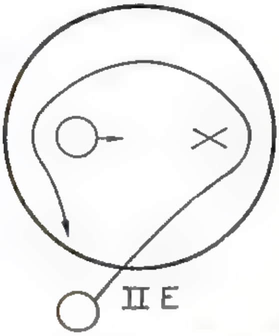


FIG. 2. Schema illustrating empathy, or projective identification. In the above diagram the observer is focusing his observations upon "how would I feel if I were in that situation."

Diagram IIE represents the perceptual set of empathy or projective identification. This is the perspective of the observer who states, "I know just how the subject feels, for if I were in his shoes, I would feel thus-and-so." While this point of view is considered unscientific, it must be recognized in order that it may be guarded against by the observer who would be consistent in the position described in Diagram II.

Even though empathy is not a valid position for scientific observation, it is a viewpoint which must be understood in terms of its significance to most laymen. Much discussion of emotional events by nonpsychologists (as well as by some psychologists) is carried on in the framework of this position. Likewise, this is the attitude which is most satisfactory when the scientist or the layman is attending a movie or play, or when he is reading stories or novels "for relaxation"!

PARTICIPANT OBSERVATION

Diagram III presents the third point of view: participant observation (13). In much of the research on emotion, this face-to-face observational position has not been differentiated satisfactorily from the two previously considered viewpoints. Yet, a brief study of the three diagrams should make it evident that this point of view, in terms of human interaction, is significantly different. Here the observer is a part of the stimulating environment of the emotional organism. In some instances, even, the observer and the X (object of emotional attitude) become identical for the emotional organism. The activities of the observer serve to enhance or reduce the intensity of the emotional behavior sequence in *S*. Likewise, there is the possibility that the observer himself may be provoked into emotional responses!

Psychologists are frequently approached by laymen requesting advice as to the means available for controlling the emotional behavior of others in such face-to-face situations. The layman as a parent may be concerned with the emotional outbursts of his children; as a husband he may be concerned with the emotional exchanges between himself and his wife; or as a teacher, policeman, or minister he may be seeking help in the person-to-person emotional encounters which are an inevitable part of the role he is playing in the social order. As a scientist, the psychologist has little professional information to offer in these situations,

although research on techniques of therapy (directive vs. nondirective, etc.) may provide tested information for the counselor in the near future. The psychologist as an artist in interpersonal relationships must draw upon his personal recollections, his common sense, his clinical experience, or upon his memory of similar situations as portrayed on the stage or in novels to offer constructive suggestions.

In psychological contexts other than those concerned with research on emotion some recognition has been given to this type of face-to-face observational situation. Moreno's (20) "role-playing" involves an approach of this nature without crystallizing the stimulative role of the observer. During World War II the OSS (25) project on personality assessment developed a test situation which is an embodiment of this point of view. The candidate, under an assumed identity, directed the activities of two young psychologists who were introduced to him as unskilled farm hands. The "farm-hand" assistants were instructed to misinterpret directions, to procrastinate, while at the same time engaging in needling conversation, all tending to throw the candidate off balance; throughout, these troublesome helpers as well as others were making note of the candidate's ability to cope with the frustrations and confusions of this unusual test situation. Such an assessment project differs from this third observational position only in that it was used for purposes other than the direct study of emotional behaviors.

Regrettably, many laboratory experiments which were set up with the intention of complete objectivity have nevertheless inadvertently slipped from the II position of objectivity to the III position of face-to-face observation. This slip may have been permitted because the experimenter considered it essential that he manipulate apparatus in coordination with *S*, or more simply that he be at close enough range to observe *S*'s responses. Some years ago the present writer published a report (18) of an experimental study of the effects upon "other hand" tension of electric shock punishment accompanying star tracing. Evidence was reported of considerable tension as measured by pressure upon a rubber bulb. A year later an advanced student was urged to collect more data on the same problem with the same equipment in the same location. The student found *little or no evidence* of tension. When the data of the unpublished study were compared in detail with the writer's earlier study, the only significant variant was that in the earlier experiment the writer as instructor used students from his own classes (with repeated assurance that participation, though required, had no bearing upon the *S*'s grade in the class), while in the later experiment the experimenter was easily identified as "only a more advanced student" who could have no influence upon the grade-giving instructor. Thus the tension reported in the published study was a function of the student-instructor relationship in this face-to-face experimental situation rather than of the electric shock.

Experiments upon behavioral phenomena other than the emotional are probably also susceptible to this face-to-face influence, particularly when prestige of the experimenter may be known to the *S*. A considerable number of check experiments would be in order to ascertain predictable distinctions between situations II and III.

LEVELS OF DESCRIPTION

In addition to the discrepancies which arise as a result of these differing observational viewpoints, there is, as men-

tioned earlier, a second basis of apparent divergence in the research studies on emotion which may be attributed to unlike levels of magnification or to unlike levels of description. Such differences in level may occur even when the fundamental point of view is identical from observer to observer.

Although the significance of the concept of levels of description is widely recognized in areas of practical measurement, it seems to have been frequently overlooked in the approach to psychological problems. Certainly it has been used rarely, if ever, in the structuring of interrelationships in the field of emotion.

A clarifying illustration of varying levels of description may be found in the English metric system for measuring length. Units are used, which vary from inches, to feet, to yards, to rods, to miles. Yet each of these units constitutes a level of description most appropriate to certain measurement activities. With small objects, such as pencils, envelopes, knives, etc., the inch is the convenient unit; with larger objects, such as lumber for building, automobiles, and furniture, the foot is the preferred unit; but still longer objects, such as a strip of cloth or a length of rope, will be dealt with in yards. Shorter distances in geographical space, such as the length of a field or a farm, are expressed in rods; while greater distances, such as those between town and town, or city and city, are expressed in miles.

A second example, that of money in economic exchange, reflects our practical acceptance of differing levels but illustrates how calibration is simplified by use of the decimal system. Our coins: cent, nickel, dime, quarter, and half dollar, are each appropriate to certain exchange activities, while our paper currency permits exchanges at various levels of higher descriptive value. Conceptually, in affairs of government, the million and billion become the convenient descriptive unit. In this context it is appropriate to point out that a person habituated to values at one level may be incapable of adequate value judgments at another level, as expressed in the folk saying, "Penny wise, pound foolish." It is likely, though perhaps difficult to demonstrate, that such preferred levels of descriptive observation peculiar to the person may be present in other areas of measurement as, for instance, in the study of emotion.

Further citation of such examples could be given in many other areas of practical human manipulation of environment. One further significant relation, however, must be indicated before we return more specifically to emotion. While we have chosen to review this descriptive level concept with hierarchies in which the units of the various levels are now calibrated in terms of each other (i.e., twelve inches to the foot), this has not always been true. Irwin (12) in a recent article in the *Scientific Monthly* pointed out that the units of the English system of measurement arose out of independent manipulative activities in different segments of the population, and were not defined in reference to each other until more general usage made their calibration necessary.

Returning now to the conceptual framework of emotion, we find ourselves in a position in which the observer (in the objective or the participant points of view) is endeavoring to use immediately experienced behavioral cues, such as gesture, facial expression, tone of voice and language, as well as situational relationships as indicators of emotional attitudes and as predictors of probable direction of ensuing motivational behavior. In certain situations, as with money in economic exchange, there are probably available a

succession of appropriate levels of description. For brevity in illustration, only the two extremes of microscopic and macroscopic observation are presented here.

Microscopic Level of Approach

If investigating at this level, the observer directs his attention toward minutiae, such as the individual contractions of the 46 or more muscles of facial expression, as did Landis (16) in the analysis of one of his photographic research studies; or, the observer tries to record successive specific acts as did Hebb's (11) associates at the Yerkes laboratory. In either instance, the observer of emotional events usually finds himself in the position of the man who would measure the distance between cities with the inch as a unit; that is, the complexity of details make the measurements difficult to comprehend unless translated into a higher-level unit.

Macroscopic Level of Approach

While the *Zeitgeist* of our scientific era has tended to make description on the macroscopic level appear to be common sense and therefore implicitly unscientific, still there is need for accurate predictive descriptions at this level, particularly for those who work with other humans in face-to-face relationships. It is essential that we anticipate and predict "the other fellow's next move" whether it be in poker, family argument, bargaining for a contract, or in phrasing a political agreement. Rarely does the "other fellow" submit to the use of a polygraph to help us ascertain his emotional uncertainties. Accordingly, most of us find it necessary to rely on the macroscopic concepts of emotion in such situations. Such concepts of emotion indicate varying degrees of uncertainty and disorganization which may be interpreted to our advantage—although, conversely, these may carry the threat of explosive disruption if pushed too far. Hence, it behooves us to go further than just detecting the emotion: we must identify its nature and predict its course. Carr (4, p. 280), in his *Psychology*, summarized this macroscopic use of the concept: "Anger is correlated with an aggressive attack against obstacles, while fear is associated with the opposite type of behavior." Love, as a similar term, is predictive in a situational reference; that is, mother love implies emotional protection of offspring; sexual love suggests amorous approach toward the loved person; while brotherly love suggests support and defense of the sibling on other than rational grounds, and the like. Anxiety suggests disruption of most coordinated life habits to a noticeable degree; terror or panic obviously implies wild irrational behaviors in escaping some real or fancied threat. This catalog could be carried on at length, although with less agreement as to predictive implication when extended to synonyms and to more infrequently used terms.

Hebb's study (11), particularly in his finding that the frankly anthropomorphic concepts proved to have more predictable use, supports our contention that most of us in our living human face-to-face environments have developed higher-order concepts for use in identifying and in predicting emotional behaviors. However, an overview of the research on the interpretation of facial expression, hand gesture, and body posture in relation to emotion would seem to indicate that as in the earlier era of independent measures of length we have not yet succeeded in calibrating our units from higher levels to the lower ones. Application of the concepts of levels of magnification and of levels of description in this context should help greatly in reconciling many

conflicting interpretations of the research data published under the topic of emotion.

SUMMARY

It is proposed in this paper that in the study of emotion, human perceptual limitations restrict the extent and the nature of observations which may be made at any time. It is suggested that this perceptual limitation extends as well to other molar concepts in psychology.

It is further proposed that much of the apparent conflict in this area could be reduced if the significant points of view were defined, and if the levels of description within each of these points of view were specified.

It is suggested that the most significantly different points of view may be specified as the (*a*) introspective, (*b*) objective, and (*c*) participant. The objective point of view occasionally carries the risk of being confused by the intrusion of projective identification. Of the three, the third (participant) has been differentiated with least clarity in the research literature on emotion.

It is indicated that several levels of description are available within each point of view, although only contrasting extremes are presented here. It is suggested further that we cannot adequately portray the whole picture until the various levels of measurement and description can be so calibrated that a given set of observations may be either integrated into higher units or differentiated into the units of a lower level.

REFERENCES

1. BARD, P. Emotion: I. The neuro-humoral basis of emotional reactions. In C. Murchison (Ed.), *The foundations of experimental psychology*. Worcester, Mass.: Clark Univer. Press, 1929. Pp. 449–487.
2. CANNON, W. B. *Bodily changes in pain, hunger, fear and rage.* New York: Appleton, 1915.
3. CANNON, W. B. The James-Lange theory of emotions: a critical examination and an alternative theory. *Amer. J. Psychol.*, 1927, 39, 106–124.
4. CARR, H. A. *Psychology, a study of mental activity.* New York: Longmans, Green and Co., 1929.
5. COLEMAN, J. C. Facial expressions of emotion. *Psychol. Monogr.*, 1949, **63**, No. 1 (Whole No. 296).
6. DASHIELL, J. F. Are there any native emotions? *Psychol. Rev.*, 1928, **35**, 319–327.
7. DEWEY, J. The theory of emotion. (1) Emotional attitudes. *Psychol. Rev.*, 1894, **1**, 553–569.
8. DUFFY, ELIZABETH. Emotion: an example of the need for reorientation in psychology. *Psychol. Rev.*, 1934, **41**, 184–198.
9. DUFFY, ELIZABETH. An explanation of emotional phenomena without the use of the concept of emotion. *J. gen. Psychol.*, 1941, **25**, 283–293.
10. DUNLAP, K. Are emotions teleological constructs? *Amer. J. Psychol.*, 1932, **44**, 572–576.
11. HEBB, D. O. Emotion in man and animal: an analysis of the intuitive processes of recognition. *Psychol. Rev.*, 1946, **53**, 88–106.
12. IRWIN, K. G. Fathoms and feet, acres and tons: an appraisal. *Scientific Monthly*, 1951, **72**, 9–17.
13. JAHODA, MARIE, DEUTSCH, M., & COOK, S. W. *Research methods in social relations.* New York: Dryden Press, 1951.
14. JAMES, W. *The principles of psychology.* Vol. II, Chap. 25. New York: Holt, 1890.
15. JAMES, W., & LANGE, C. G. *The emotions.* Baltimore: Williams and Wilkins, 1922.
16. LANDIS, C. Studies of emotional reaction: II. General behavior and facial expression. *J. comp. Psychol.*, 1924, **4**, 447–509.
17. LANDIS, C. Emotion: II. The expressions of emotion. In C. Murchison (Ed.), *The foundations of experimental psychology.* Worcester, Mass.: Clark Univer. Press, 1929. Pp. 488–523.
18. MCTEER, W. Changes in grip tension following electric shock in mirror tracing. *J. exp. Psychol.*, 1933, **16**, 735–742.
19. MEYER, M. Emotion: that whale among the fishes. *Psychol. Rev.*, 1933, **40**, 292–300.

20. MORENO, J. L. *Who shall survive.* Washington: Nervous & Mental Disease Pub. Co., 1934.

21. MUNN, N. L. Feeling and emotion in everyday life. Chap. 15, pp. 365–369, in *Psychology, the fundamentals of human adjustment.* (2nd Ed.) Boston: Houghton Mifflin, 1951.

22. SHERRINGTON, C. S. Experiments on the value of vascular and visceral factors for the genesis of emotion. *Proc. royal Soc. of London,* 1900, 66, 390–403.

23. WEISS, A. P. Feeling and emotion as forms of behavior. In M. L. Reymert (Ed.), *Wittenberg symposium on feeling and emotion.* Worcester, Mass.: Clark Univer. Press, 1928.

24. YOUNG, P. T. *Emotion in man and animal.* New York: Wiley, 1943.

25. OSS ASSESSMENT STAFF. *Assessment of men.* New York: Rinehart, 1948.

[MS. received May 23, 1952]

Psychological Review
Vol. 60, No. 3, 1953

DO INCORRECTLY PERCEIVED TACHISTOSCOPIC STIMULI CONVEY SOME INFORMATION?

PETER D. BRICKER AND A. CHAPANIS
The Johns Hopkins University

In a recent experiment on "subception," Lazarus and McCleary concluded that "a process of discrimination can operate prior to conscious recognition and in the absence of the possibility of the correct verbal report" (3, p. 116). The process with which they were concerned was autonomic discrimination among visual stimuli. They first established a conditioned GSR to 5 of 10 tachistoscopically presented nonsense syllables by pairing them with electric shock, and then measured the GSR for subsequent presentations of the 10 stimuli without shock. For only those stimuli which subjects failed to identify in a single verbal report, the average GSR following shock syllables was significantly greater than the average GSR following nonshock syllables. The gist of their results appears in Fig. 1.

Since the Lazarus-McCleary experiment was undertaken to provide support for a number of recent articles having to do with the influence of need on perception, their discussion of the problem is oriented in this direction. We have no quarrel with their procedures and data, but we regard as unfortunate the implication that they have discovered evidence for some sort of "unconscious determination of behavior" (3, p. 121) operating when "recognition . . . is impossible" (3, p. 114). We prefer to interpret these data as suggestive of a rather more obvious principle which is important for many psychological experiments: Even when an S's first verbal response to a stimulus is wrong, the stimulus may still have conveyed useful information to him. If this is so, it should be possible to demonstrate this transfer of information by some sort of verbal response in the absence of shocks, needs, or other strong emotional provocation.

Statement of the Problem

The study reported here is an attempt to answer this question: Is it possible to show by some kind of verbal response that a stimulus has conveyed useful information, even when the stimulus has been incorrectly identified on the first trial? As our verbal response, we have used the number of additional guesses, following the initial wrong response, necessary to name the stimulus correctly. If an *S* needs fewer additional guesses than are to be expected by chance, we may infer that the likelihood of a correct response has been increased by the stimulus, i.e., that the stimulus has conveyed some information. A secondary aim of this study is to get some insight into the factors which operate to determine the *S*'s verbal behavior in a guessing task such as this.

Method

Subjects

Ten male undergraduates served as *S*s. None had had more than an elementary course in psychology.

Apparatus

The *S* sat seven feet from a milk-glass screen on which the stimuli were projected from a tachistoscope. A between-lens shutter and a variac in the bulb circuit of the tachistoscope permitted variations in both exposure speed and intensity. The stimulus words were five-letter double nonsense syl-

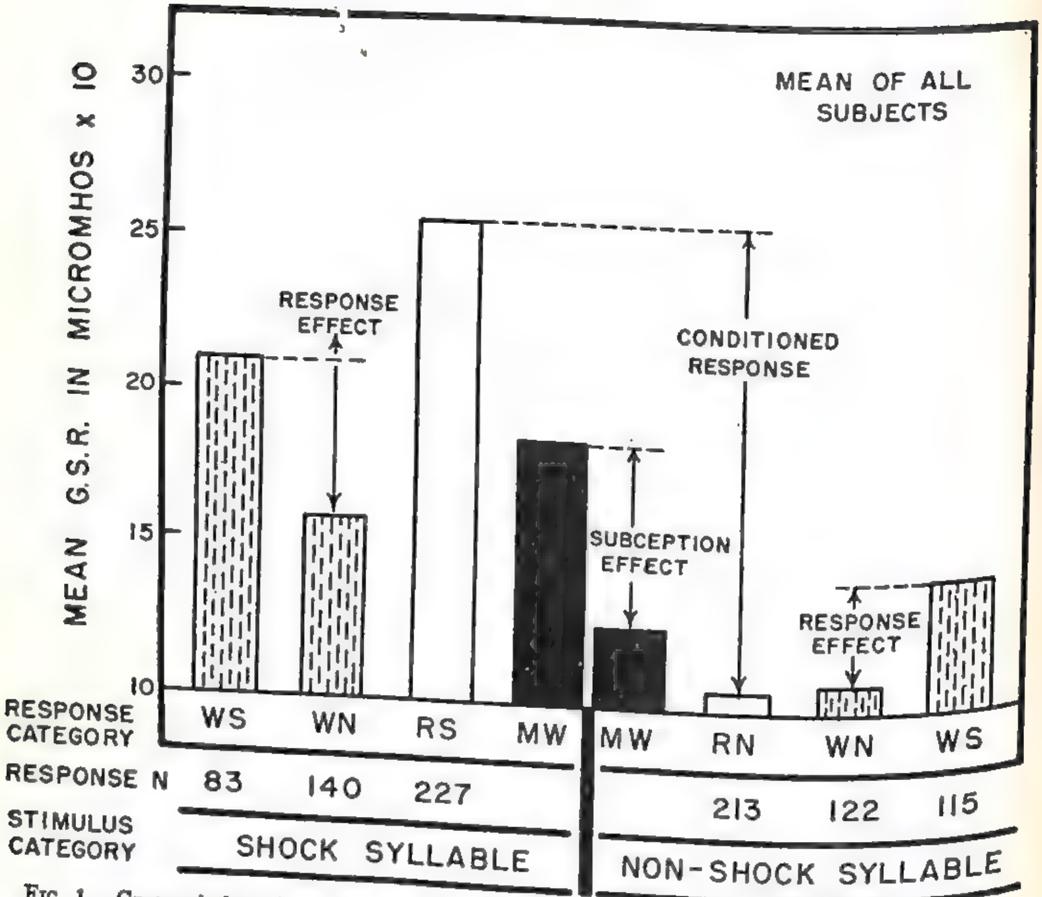


FIG. 1. Grouped data from the study by Lazarus and McCleary (3). The heights of the bars indicate the average size of the GSR over nine *S*s for each of the classifications of stimulus and response. The Response Category labels may be interpreted with the following guide: W indicates that the response was wrong, R that it was right; S indicates that the response was one of the syllables which had previously been associated with shock, N that it was a nonshock syllable; and M denotes the arithmetic mean. Thus, the bars labeled MW are the mean GSR's for all wrong responses following each type of stimulus syllable. Each MW bar is also the average for the two striped bars in its stimulus category.

lables, or paralogs, printed on slides in capital letters. Eight of these stimuli (hereafter called the List stimuli) were listed on cards with which the *S* was supplied, and five of them (hereafter called the Nonlist stimuli) were known only to *E*. Both groups of stimuli are shown in Table 1. The *S* was provided with eight cards, on each of which was a different random arrangement of the same eight List stimuli.

Procedure

Preliminary series. The *S* was told that he was taking part in an experiment on the training of perception, and was never informed of the Nonlist stimuli. He was also told to respond after each exposure, even if he had no idea what the stimulus was. During the preliminary series, the List stimuli were presented in random order, and *S* was allowed one guess for each stimulus. The *S* used his cards as guides in making his responses, putting the card he had just used on the bottom of the pack after each guess. This was a precautionary measure to prevent the *S*'s establishing order preferences in the use of the syllables. As the preliminary series progressed, *E* adjusted the shutter speed

TABLE 1

LIST AND NONLIST STIMULUS WORDS*

List Stimuli	Nonlist Stimuli
GOKEM	LAJYV
TAROP	NIGAT
LATUK	RUNIL
SIJUD	VECYD
HEZUW	YUZYJ
FEXAD	
MYZEG	
CEFIJ	

* The List stimulus words are those the subject knew about and tried to guess. Subjects were not aware that the Nonlist stimulus words were also shown during the experimental trials.

and variac setting until S consistently got half or less of the stimuli correct. The speed and illumination arrived at in this way were used throughout the experimental run. This session also allowed S to become familiar with the stimuli and to achieve an asymptotic level of proficiency in recognizing them.

Test series. After a five-minute rest period, the experimental series of 120 presentations was begun. In this series, the 13 stimuli were arranged in random order, with the restriction that each List stimulus appeared ten times and each Nonlist stimulus eight times. One further restriction, which will be explained later, was that each List stimulus was designated as the correct response for each Nonlist stimulus only once. The E informed S that he must keep guessing after each stimulus, without seeing it again, until E said "right." The S again used his cards as guides, and changed cards only after completing a series of responses to one stimulus. The S wrote down each guess as he reported it verbally to E so that he would not repeat any guess which was wrong. The E said "wrong" or "right" after each guess.

Nonlist stimuli were included in the experimental series in order to obtain from each S a distribution of the number of additional guesses needed to identify the List words when they were not presented to him as stimuli. Since perceptual conditions were difficult, it was safe to assume, as later questioning of the Ss proved, that the S would not discover that stimuli other than those on his cards were being presented. Hence when a Nonlist stimulus was presented, the S began making guesses from his group of eight List stimuli. For each presentation of a Nonlist stimulus, a List word was selected arbitrarily, within the limits described in the preceding paragraph, as the correct response. Thus all conditions which pertained when S was trying to identify a relevant stimulus were the same, except that the stimulus could not contain any useful information. This procedure is effectively the same as asking the S to make 40 series of guesses at the List word of which E is thinking, when E thinks of each of the eight words at random five times during the sequence of 40 trials.

RESULTS

List vs. Nonlist stimuli. In answering our primary question, we worked only with those stimuli to which the initial response was wrong. The mean number of additional guesses necessary to make the correct response after (a) List stimuli and (b) Nonlist stimuli were computed for each S (see Fig. 2). The difference between these means was computed for each S and entered into a distribution of differences. The mean number of additional guesses for List stimuli for all Ss is shown in Fig. 2 as a solid line, and that for Nonlist stimuli is shown as a dotted line. Since the probability that the mean difference of .63 occurred by chance is less than .001, we conclude that significantly fewer additional guesses were needed to identify incorrectly-perceived stimuli than to guess names for stimuli when the names had no meaningful connection with the stimuli. In other words, the List stimuli conveyed some useful information to S, i.e., increased the likelihood of his making a correct verbal response.

To determine whether the guessing behavior following Nonlist stimuli was different from chance, we made use of all 40 guesses made by each S after the

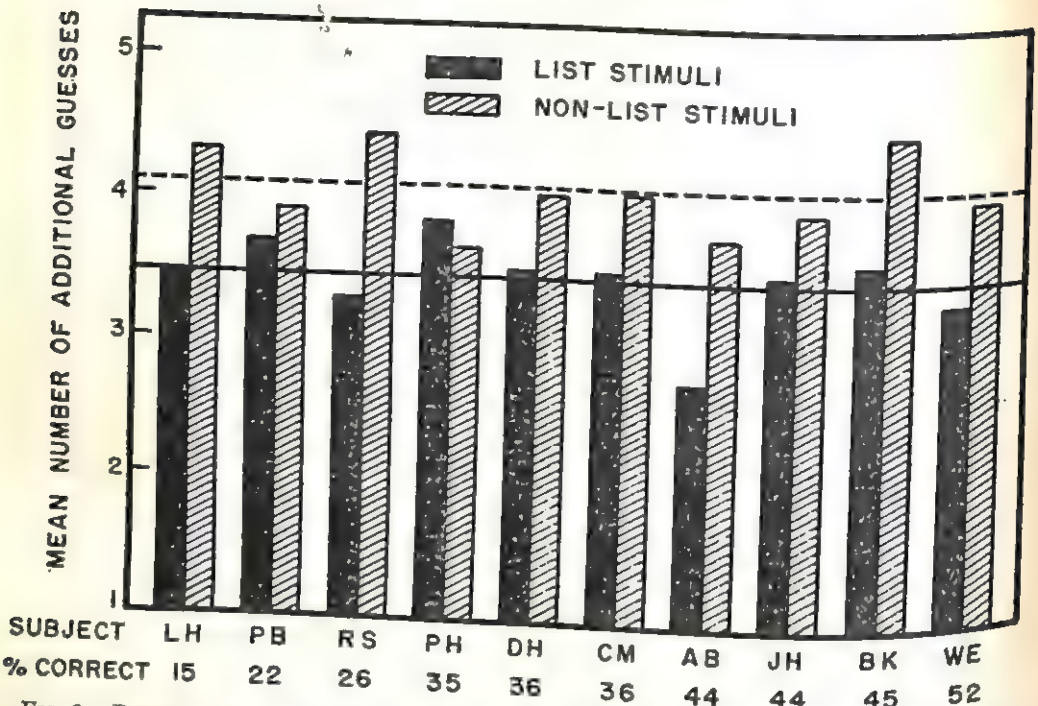


FIG. 2. Results of the test series for 10 *Ss*. The black bars show the mean number of additional guesses needed to identify List stimuli after an initial wrong response; the striped bars show the mean number of additional guesses needed to make the correct response after Nonlist stimuli. The solid horizontal line is the mean of the black bars; the dashed line the mean of the striped bars.

Nonlist stimuli. The mean number of guesses necessary to make the correct response following Nonlist stimuli was 4.58, while the mean to be expected by chance was 4.50.[1] The difference between the mean and the chance value is not significant, so we conclude that *Ss* were guessing at random when they had to identify List words following Nonlist stimuli.

The various *Ss* were able to identify from 15 per cent to 52 per cent of the List stimuli correctly on the first response, as shown in the bottom line of Fig. 2. This figure also shows that the magnitude of the difference between the two mean numbers of additional guesses for each *S* is not related in any obvious way to the percentage of List stimuli which he was able to identify correctly on the first response.

Word preferences. The number of times each word was used as a first response was compared with expected frequencies for each *S*. This was done first for responses to List stimuli alone, and then for all 120 first responses. The combined chi squares for both of these conditions were significant beyond the .001 level, indicating that *Ss* respond more frequently with some words than with others. When the words are ranked for each *S* according to preference, however, the average intercorrelation by ranks among *Ss* is only 0.234.[2] This means that there was little con-

[1] The apparent discrepancy between this statement and the data of Fig. 2 is easily resolved by noting that Fig. 2 shows the mean number of *additional* guesses following initially wrong responses.

[2] The average rank intercorrelation was computed by the technique derived by Peters and Van Voorhis (5, pp. 200–201).

sistency among *S*s as to which words were preferred or nonpreferred. However, the preference rankings of two of the words were quite consistent from subject to subject. LATUK was ranked first or second by 9 out of the 10 *S*s, and third by the remaining one. SIJUD, on the other hand, was the least-preferred first response. For six *S*s it ranked seventh or eighth, and its other ranks were 6, 5, 3, and 2. It is quite possible that LATUK occurred more frequently as a first response partly because it was more easily recognized than other words by most of the *S*s.

Word-sequence preferences. Tabulations of first-order dependencies in sequences of two or more responses to one stimulus indicate preferences for certain sequences of words. Although our data are too few for many detailed comparisons, there are many instances in which words with similar elements (letters) follow each other. The data for one *S*, for example, show that the two permutations of SIJUD and FEXAD account for 44 per cent of the instances in which either is followed by another word. Another *S* said GOKEM more than half the time after MYZEG and HEZUW, while 80 per cent of the responses following GOKEM were one of the latter two words. Still another followed LATUK with TAROP 40 per cent of the time. CEFIJ and SIJUD seem to occur in pairs for many of the *S*s. Some of the similarities which do not seem obvious when viewing conditions are normal might be clearer if we knew what elements of the words stand out under difficult tachistoscopic conditions.

Word legibility. Additional computations suggest that there are differences in ease of recognition (accuracy) among the words, and that the relative position of the words in this respect also varies from *S* to *S*. There were, however, too few responses per word to determine the relative accuracy of the words or to correlate word preference with accuracy within *S*s.

DISCUSSION

Our results agree with the findings of Lazarus and McCleary (3) in providing an affirmative answer to the question posed in this paper: Even when *S* gives an incorrect verbal response to a stimulus, the stimulus may still have conveyed some useful information. Lazarus and McCleary demonstrated this by a conditioning technique. We have shown the same thing by using verbal responses. Following an initial wrong response, the number of additional guesses necessary to name the stimulus correctly is fewer than would be expected by chance.

Unlike Lazarus and McCleary, however, we look for more ordinary interpretations of our results. We believe that in any situation where perceptual conditions are difficult, *S* may receive meaningful cues from the stimuli and still make wrong verbal responses. The cues have the effect of narrowing the possible responses to a few, or of establishing for *S* a class or group of stimuli of which he is certain that the stimulus just presented was, or was not, a member. For a good illustration of how this happens in a perceptual situation somewhat different from ours see Craik and Vernon (1). There remains the question: What are some of the factors that influence the *S*'s verbal response in a guessing task such as this?

The work of a number of *E*s makes it clear that there are many determinants of the response in a psychophysical judgment. The sensory threshold is probably the most important factor, but there are others which have pronounced effects. Preston, for example, showed that *S*s tend not to repeat the judgment just previously made (6). In

addition he (see also Schafer [7]) reports pronounced contrast effects in successive judgments. The probability is better than .5 that a near-threshold stimulus will be reported as not sensed if it follows a stimulus which was clearly above threshold. Subjective preferences for numbers and sequences of numbers have been found by Yule (9) and Chapanis.[3] And so on.

In this experiment there are several important factors that seem to have partially determined the subject's verbal response. These are:

1. *Word legibility.* This point is so obvious as scarcely to require comment. Some words are more legible than others under any given set of exposure conditions. Note, however, that this factor can operate in a negative, as well as a positive way. When a stimulus has been seen indistinctly, the *S* may tend to avoid using the names of words which are highly legible.

2. *Letter legibility.* Even more important for this experiment is the factor of letter legibility. Some letters are more legible than others under any given set of exposure conditions. (See, for example, Tinker [8].) In this study introspective evidence strongly supports the importance of this factor, as do the data on word-sequence preferences. If only the letter J was seen clearly—as sometimes happened during a trial series of exposures for the junior author—the *S* first guessed either SIJUD or CEFIJ, and, if that response was wrong, followed it with the other alternative. Sometimes it was possible to see rather distinctly that the second letter was an A—thus reducing the range of possible choices to only two.

3. *Word preferences.* Over and above the factors of word or letter legibility, there is some evidence that, in the absence of other cues, *S*s consistently preferred to use some words as responses and rather consistently avoided others.

4. *Word-sequence preferences.* Not only are certain words preferred over others, but there is good evidence that *S*s tend to use words in certain sequences. We have shown previously that many of these sequences appear to depend not upon subjective preferences, but upon elements common to the component words. Whether all such sequential preferences are the result of common recognizable elements cannot be determined from our data. Many similar elements are evident in pairs such as SIJUD and FEXAD, CEFIJ and SIJUD, LATUK and TAROP, and so on, but with our data we can only point to these sequence preferences without explaining them. We may recall, however, that even nonsense syllables have some associative value, and that in his investigation of the Zenith radio experiments in telepathy Goodfellow (2) found very striking preferences for certain sequences of meaningful, but unrelated, words.

Now for the question: Can factors like these account for the effects found in the Lazarus-McCleary experiment? We think so. In the Lazarus-McCleary data (see Fig. 1) mean GSR's are given for four classes of wrong responses. The largest average GSR occurred after shock stimuli which the *S*s identified as other shock stimuli, i.e., the *S*s had placed the shock stimuli in the correct class. The smallest average GSR occurred following nonshock stimuli which the *S*s identified as other nonshock stimuli. Intermediate and nearly equal mean GSR's were found after shock stimuli which the *S* called nonshock words, and vice versa. In this connection, note too that the average response effect is almost as large as the average "subception" effect. This means that the size of the GSR is largely, but

[3] Chapanis, A. Unpublished data.

not entirely, dependent on whether *S* thought (as evidenced by his first verbal response) that the stimulus was a shock word, or nonshock word.

Now let us make some hypotheses about what the *S*s might have been doing in the Lazarus-McCleary experiment. Suppose, for example, that we had conditioned SIJUD and CEFIJ to shock and had used FEXAD, LATUK, and TAROP as nonshock syllables. Suppose further that a word is presented and that the *S* manages only to recognize an A somewhere in the word. We would expect his verbal response to be either LATUK, TAROP, or FEXAD, but in any case we would expect a small GSR, because none of the plausible responses is a shock word. Suppose, on the other hand, that *S* recognized, or thought he recognized, only an IJ somewhere in the word. His verbal response will be either SIJUD or CEFIJ, but *right or wrong*, we would expect a large GSR because the only plausible responses are shock words.[4]

Finally, suppose that SIJUD is presented and *S* recognizes only the final D. The plausible responses are FEXAD, which is wrong, and a nonshock word, or SIJUD, which is correct and a shock word. In this instance the first verbal response is misleading because the *S* may be perfectly aware that the stimulus may have been either a shock or nonshock word. Thus, we would expect the GSR to be moderately large because the plausible responses include both a shock and nonshock word and, so far as the *S* is concerned, it could be one or the other. If *S* gives the wrong verbal response, his GSR is entered into the WN category for shock stimuli in Fig. 1. Note, however, that exactly the same kind of reasoning applies to the GSR's in the WS category for nonshock stimuli in Fig. 1. Suppose that FEXAD, a nonshock syllable, is presented and *S* recognizes only the final D. A GSR of intermediate size would again be expected since the plausible responses still include a shock and nonshock word. If the *S* guesses wrong, his GSR is entered into the WS category for nonshock stimuli. Such cues, it seems to us, could account entirely for the relative heights of the four striped bars in Fig. 1 and, since the solid bars are averages of the striped ones, for the subception effect as well.

Lazarus and McCleary point out, and this is also evident from Fig. 1, that when the *S*s were wrong, they showed no tendency to categorize correctly. That is, when a shock syllable was presented and *S* guessed wrong, he was not more likely to guess some other shock word. Similarly, when a nonshock word was presented and *S* guessed wrong, he was not more likely to guess some other nonshock word. Indeed, there is some evidence that the reverse occurred. This suggests that the *S*s were categorizing on some basis other than the previous shock, or nonshock, association of the word. Thus, this observation lends further support to our interpretation that *S*s were categorizing on the basis of common recognizable elements among the words.

In this connection, it is important to point out that the Lazarus-McCleary syllables appear to have more common elements than ours. They used YILIM, ZIFIL, JIVID, JEJIC, GAHIW, GEXAX, YUVUF, ZEWUH, VAVUK, and VECYD. Note that three words have the letter I, an easily recognized letter, occupying the second and fourth positions. Two other syllables have an I in the fourth place. Two words each start with Y, J, G, Z, and V. Three words have a U in the fourth position, and so on. The pres-

[4] For evidence that conditioned responses can be evoked by elements of complex stimuli, see Pavlov (4, p. 142 f.), for example.

ence of such a great number of common elements among their syllables makes our explanation of the data more likely. We strongly suspect that cues such as these carry the information which enables us to "subceive."

SUMMARY

Ten *S*s were shown nonsense syllables under conditions such that they could recognize half or less of the syllables correctly on the first trial. For only those stimuli which subjects failed to identify on the first trial, the mean number of additional guesses necessary to name the stimulus correctly was compared with the mean of a distribution of random guesses following stimuli which were not available to the *S*s as responses.

1. Significantly fewer additional guesses were needed to identify incorrectly perceived stimuli than to make the correct response in a series of random guesses. Tests confirm the randomness of the distribution of guesses for stimuli about which the *S*s were uninformed.

2. Subjects were able to guess from 15 to 52 per cent of the stimuli correctly on the first trial. The superiority of guessing incorrectly perceived stimuli over random guessing seems to hold for the entire range.

3. Subjects exhibited preferences for certain words as first responses. They were not consistent among themselves as to which words were preferred.

4. Certain sequences of words seemed to occur more often than others. Some of these sequences are composed of words which have similar elements. Recognition of these elements may be the basis for such information as is conveyed, and may provide the explanation for the results of recent experiments on "subception."

REFERENCES

1. CRAIK, K. J. W., & VERNON, M. D. Perception during dark adaptation. *Brit. J. Psychol.*, 1942, **32**, 206–230.
2. GOODFELLOW, L. D. A psychological interpretation of the results of the Zenith radio experiments in telepathy. *J. exp. Psychol.*, 1938, **23**, 601–632.
3. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: a study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.
4. PAVLOV, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep.) London: Oxford Univer. Press, 1927.
5. PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.
6. PRESTON, M. G. Contrast effects and the psychophysical judgments. *Amer. J. Psychol.*, 1936, **48**, 389–402.
7. SCHAFER, T. H. Influence of the preceding item in measurements of the noise-masked threshold by a modified constant method. *J. exp. Psychol.*, 1950, **40**, 365–371.
8. TINKER, M. A. The relative legibility of the letters, the digits, and of certain mathematical signs. *J. gen. Psychol.*, 1928, **1**, 472–496.
9. YULE, G. U. On reading a scale. *J. roy. statist. Soc.*, 1927, **90**, 570–587.

[MS. received April 21, 1952]

Psychological Review
Vol. 60, No. 3, 1953

A MORE RIGOROUS THEORETICAL LANGUAGE

JACK L. MAATSCH AND RICHARD A. BEHAN[1]

Michigan State College

The purpose of the present paper is to discuss the relevance of an explicit metalanguage in contemporary psychological theorizing. We shall attempt to show the importance of specifying the rules concerning meaning and denotation of terms used in theory construction, and to set forth rules for the admission of constructs as adequate for theory construction.

To illustrate some of the criticisms which we shall make, we have chosen examples of unfortunate usage in modern psychological theorizing. We do not, by this device, wish to call attention to specific theories as inadequate. Rather, we feel that these unfortunate usages are typical of much of psychological theorizing.

Part I

1. *Theory as a form of language.* Every theory is a system of language, and as such, shares certain characteristics of all languages. The first distinction we make is between *language* and *metalanguage*.

When we use the word *language*, we refer to a system of habits or activities of human beings which serves the purpose of communication and of coordination of activities between the members of a group. By the word *metalanguage* we refer to a system of habits or activities which serves the purpose of discussing the use of the language. In other words, a metalanguage is a language which has as its subject matter a language proper. Thus every language has two parts: (*a*) the metalanguage, which contains rules for the use of the language proper; and (*b*) the language proper, which is used for the purpose of communication. The metalanguage in turn is usually divided into three parts: syntactics, semantics, and pragmatics (4).

The syntax of a metalanguage contains (*a*) an enumeration and listing of the signs of the theory; (*b*) rules which determine when an expression is significant, and when a significant expression is a part of the theory; (*c*) rules which determine when a given significant statement is a deduction from other statements of the theory. All of these tasks which belong to syntax may be accomplished without reference to the meanings of the signs and expressions of the theory (13).

The semantics of the metalanguage is concerned with problems of denotation, truth and falsehood, and meanings of the signs of the theory. The rules of the semantics determine what is denoted by the terms of a theory, when a statement in a theory is true or when it is false (4).

The pragmatics of a metatheory is concerned with the way in which the terms of the theory are conventionally used. The rules of pragmatics contain statements which describe the uses of the various terms in the language (4, 13).

Examples of syntax, semantics, and pragmatics taken from the English language are: (*a*) Syntax, "Every complete sentence contains a subject and a

[1] The authors wish to express their gratitude to their friends and professors who so kindly criticized the first draft of this paper. We especially owe thanks to Professors Henry S. Leonard and Lewis Zerby of the department of philosophy, and to Professors M. Ray Denny and Donald Johnson of the department of psychology.

verb," "An interrogative sentence is ended with a question mark"; (*b*) Semantics, "The word 'mother' is properly used when a female person stands in 'gave-birth-to' relation to another person," "The word 'trial' denotes a sequence of events in which some person is tested to determine innocence or guilt of some crime"; (*c*) Pragmatics, "The word 'ocean' is used to refer to a large body of salt water which covers a great part of the earth," "The word 'sea' is used to refer to a large body of salt water smaller than an ocean and usually surrounded by land," "The word 'lake' is used to refer to a body of fresh water usually surrounded by land."

The notion of a metalanguage has its analogy in psychological theory in the "point of view," or the "school of thought," or the "frame of reference" of the theorist as he approaches a particular subject-matter area. These terms designate particular approaches to the subject matter of psychology, e.g., Behaviorism, Purposive Behaviorism, Dynamic Psychology, etc. The different approaches as they indicate a particular semantics and a particular pragmatics represent different informal metalanguages, which serve to control the use of their distinctive terminology.

It is important to note that while all languages contain syntactic, semantic, and pragmatic rules, for any given language these rules may not be explicit. If these rules are not made explicit, it is possible for the theorist to construct statements which are not only ambiguous but are meaningless for the particular theory under consideration.

2. *Conversational language as opposed to theoretical language.* By the term *conversational language*, we shall understand any one of the vulgar languages, e.g., English, French, German, etc.

By the term *theoretical language*, we shall understand a particular language designed specifically for the purpose of making unequivocal assertions about a given subject matter. In this distinction we follow Carnap (4).

The distinctive differences between conversational and theoretical languages, as we would like to use the terms, arise from the specificity of the metalanguage associated with a theoretical language, as compared with the relative lack of specificity of the metalanguage associated with a conversational language.

Contemporary psychology in using a conversational language with its inexplicit and ill-defined metalanguage creates for itself many purely linguistic problems. First, in the area of semantics and pragmatics we find ambiguity of meaning associated with particular symbols. Ambiguity arises in a conversational language because of the multitude of contextually determined meanings a symbol may have.

It is easy to slip, unnoticed, from one meaning of a word to another—to use two different expressions with the same symbol in the same discourse. As an example of the use of one symbol with more than one meaning, consider the following passages from Mowrer and Lamoreaux (11). "But fear will continue to be present between trials; and when the rat tries likewise to deal with this fear by leaping, nothing happens—the fear is not reduced" (11, p. 198). And again, "Since it is the situation as a whole which has become the conditioned stimulus for the fear reaction, any response which will remove the rat from the situation will be powerfully reinforced by fear reduction. In this way the leaping can be differentially strengthened and made to rise rapidly in the rat's hierarchy of responses to the acquired drive of fear" (11, p. 197). In another place, "the intermediate conditioned response-drive of fear."

The word "fear" is used in the above quotations with the following different meanings: (*a*) as a stimulus; (*b*) as something with which the rat attempts to deal, i.e., something experienced by the rat; (*c*) as a stimulus-produced drive; (*d*) as a response-produced drive; (*e*) as an emotional response; (*f*) as an acquirable drive. Furthermore, it is often difficult to determine from the context of usage just which one of the many meanings is meant in any given instance.

It would seem that if we are to have some degree of clarity in a theoretical language, it would be necessary to make explicit the relationship between a symbol and its meaning.

First we would like to distinguish between the terms *symbol* and *expression* by adopting a convention proposed by Lewis (8, pp. 73–74):

> . . . Linguistic signs are verbal symbols. A *verbal symbol* is a recognizable pattern of marks or of sounds used for purposes of expression and communication. . . . Two marks, or two sounds, having the same recognizable pattern, are two *instances* of the same symbol, not two different symbols. . . . A *linguistic expression* is formed or determined by the association of a symbol with a fixed meaning. . . . If in two cases, the symbol is the same but the meanings are different, then there are two expressions; not one. Also, if in two cases the meaning is the same but the symbols are different, then there are two expressions; not one. But if in two cases . . . the symbol is the same and the meaning is the same, then there are two *instances* of the expression, but only one expression.

To be consistent with our adopted usage of the relation of a symbol and its meaning we would consider that the above uses of "fear" would represent six different expressions. The writers quoted distinguished two, "S_f" and "R_f," namely (*a*) and (*e*) above, respectively; and used the word "fear" indiscriminately with the other four. As these examples indicate, it is imperative that any theoretical language system should be so constructed that the meanings which are carried by any symbol should be unmistakable. The above examples also testify to the inadequacy of the conversational language as a vehicle for scientific theory. This inadequacy of the conversational language as a medium for psychological theory lies in the fact that the rules of the metalanguage are not specified with sufficient rigor.

Secondly, a syntactical problem concerning conversational language—the fact that the conversational language does not lend itself to deductive procedures—results in the theorist's attempting to make deductions on the basis of the *meanings of expressions* contained in the statements of the theory, instead of on the basis of the *form of the statements* contained in the theory. For a discussion of these problems see Woodger (13). Owing to the fact that the structure of the conversational language precludes the possibility of using valid deductive procedures, the theorist attempts to base deductions on meanings which may be inferred from the context and which seem to be the same. It is this similarity of meaning which leads to the illusion that such and such statements are valid consequences of such and such other statements. It is indeed interesting that few of our present-day theorists publish their deductive arguments in any rigorous symbolic form.

Part II

1. *Introduction*. Psychological theorists have tended to make distinctions between different types of constructs which might be used in developing a theoretical language (8, 9). The following four notions have been used by psychological theorists as a basis for

differentiating between different kinds of constructs: operational definition, reality status, surplus meaning, and specificity of formulation.

We intend to analyze each of these notions separately and to show that they cannot be used as a basis for distinguishing between different kinds of constructs. Finally it will be our task in this section to formulate a set of rules which will determine the characteristics of constructs as special types of symbols that are employed in a theoretical language.

2. *Operationism and meaning.* We should like to discuss briefly two familiar notions concerning the definition of constructs. These notions are the meaning and the significance of a construct. Any construct, if it is to be of use in an empirical theory, must have meaning and must be significant to the science. To have meaning it must be operationally defined; to be significant—that is to possess explanatory and predictive value—it must be explicitly related to other constructs in the theory (1). These are to be considered the minimal criteria for the admission of constructs.

The meaning of a construct is the meaning given by the definiens of its definition and nothing more (3; 8, pp. 134–135). The use of a description of operations and their effects as the meaning of constructs serves at least three purposes. First, the meaning is specified in terms that have unambiguous denotations. Secondly, the construct is exposed to the possibility of experimental manipulation with the result that the theory thus specifies what operations must be used and what effects must be observed before the theorist may claim that the experimental results are consequences of the operation of the constructs in the situation. Finally, the definition presents adequate criteria upon which to assert or deny the presence of the construct for explanatory purposes.

These three characteristics of the operational definition would seem to be absolutely essential for the admission of any construct into an empirical theory. If the construct is to be useful in a predictive sense, then it must bear explicit relationships to other constructs in the theory, since the prediction and explanation of phenomena is the proper task of theory. Thus, demonstration of the construct by achieving certain results with a given set of operations does not alone constitute adequate grounds for the admission of a construct into an empirical theory. The specification of the operations and acceptable results merely define the construct in question.

The present-day theoretician—insofar as Boring speaks for him—is apparently not interested in the notions of operationism and formalization. Boring says,

> The reduction of concepts to their operations turned out to be dull business. No one wants to trouble with it when there is no special need. The reduction takes thought and study, and they take time, and there may be little or no gain. A still more rigorous language is furnished by symbolic logic, but no one wants to reduce James's *Principles* to a set of postulates and conclusions after the manner of Hull in his most exact moments. The operational technique seems to have become something to use when the user thinks he can get somewhere with it (2, p. 658).

We would differ with the position expressed in the above quotation in some particulars. First, the reduction of concepts may be dull business, but this is equivalent to saying that the giving of exact empirical meaning to one's concepts is dull business. Whether it is or is not a dull business, it is important that empirical concepts should have exact empirical meaning. This is achieved

by the use of the operational definition. The same remarks apply with respect to the second and third lines of the quotation. There may be no special need to give our empirical constructs meaning, and it is true that little may be gained by so doing—but the little that will be gained is the possibility of eventually constructing theories which will be testable. Second, the fact that no one wants to reduce James's *Principles* to statements in symbolic logic is not an argument against the use of symbolic logic in scientific psychology. It is an extremely interesting paradox that many psychologists look forward to the day when psychological theory will be couched in mathematical terms, yet symbolic logic, a more versatile language and one which makes few assumptions about the data, has received little attention. If we are going to express the relations which must obtain between constructs in a form which will allow a rigorous deductive methodology, then we shall be forced to the use of symbolic logic in one form or another.

3. *Reality status.* There has been some tendency in the past few years to advocate an appeal to the reality of constructs, as opposed to operational definition, as a basis for their admission in theory construction. The procedure is to assume that the constructs are "... actually existing structures which might eventually be described by direct experimentation. . . . Because it is assumed that these hypothetical constructs exist, and because of the intrinsic properties that they are assumed to have, the correlations between experimental conditions and results are . . . seen as necessary correlations, as inevitable consequences of the functioning of these hypothetical constructs" (7, p. 283). Or again, "genuine hypothetical constructs cannot be isomorphically related to a system of neural events, but must be a system of neural events . . ." (7, p. 284).

It is seen that the emphasis is on the realness, the existence, or the genuineness of the construct. For example, the realness of *Dynamic Systems* (7) is derivative from the assumption of the realness of neural events. However, the meaning of the theory of neural events is given, by the physiologist, in terms of operations, or by appeals to chemistry and physics. What is meaningful to the physicist is again dependent upon operations, but what is real to him depends on his particular metaphysics. For example, Newton began his *Principia* with *Space, Time, Matter.* These were the real things of which the world was composed. Other entities were known only in terms of these three. Modern quantum physics is based on two assumptions (as per Einstein); these are space-time (field) and matter (5). Still other theoretical physicists have constructed other metaphysical systems which do not mention the word "matter," but which are satisfactory for the development of the concepts of the science. See especially Whitehead (12).

It seems that what is considered real in one science is given meaning in a presupposed science in terms of operations, and presuppositions about the subject matters of still other sciences. In the final analysis, meaning is given in any science in terms of operational definition—only after the theorist has arbitrarily decided what are the real things to which he wishes to reduce his concepts. We would conclude that Hull's $_{S}E_{R}$ (6, p. 242) is just as real as Krech's Dynamic Systems. Indeed, more real—Hull's $_{S}E_{R}$ has been, at least in part, operationally defined and hence has empirical meaning.

Individuals who distinguish types of constructs on the basis of reality status seem to forget that theories are sets of

communicable signs. The operational definition is simply a device to transfer sense meanings from one group of symbols to another. The meaning which is conveyed by the definiens—the description of events in the real world—is transferred to the definiendum, the symbol which denotes the construct in question.

We feel that each theorist has a right to make any assumptions whatsoever about what is real and existent in the world. These metaphysical assumptions are prior to the particular empirical system which is to be constructed. For the empirical scientist, since the constructs are operationally defined in terms of "reality," the only criterion of an adequate metaphysics is the predictive efficiency and the explanatory power of the resulting empirical theory.

By *predictive efficiency and explanatory power* we mean one and the same thing—the ability to deduce statements descriptive of phenomena as valid consequences of statements in the theory. The distinction between prediction and explanation lies not in things that the theorist does, but rather in the intent of the theorist. In both cases the theorist does the same thing—he attempts to deduce statements which describe empirical phenomena as valid consequences of a theory. In the case of *prediction* the intent of the theorist is to *test* the theory. In the case of *explanation*, the intent of the theorist is to *understand* the phenomenon. Therefore to assert that a theory is to aid in prediction and explanation is to say only one thing, not two. Furthermore, by the phrase "deduce as valid consequences" we mean deduction which turns on the form of the statements, and not deduction which turns on the fact that two sets of statements contain words (expressions) which may be construed in the same way.

4. *Surplus meaning and specificity of formulation.* By surplus meaning we refer to the notion that a construct may have meaning beyond that stated in its definition or in the laws which relate it to other constructs in a theory (9). The notion of the surplus meaning of constructs arises from at least three sources: first, the assumption of the reality status of constructs; second, the use of discussion instead of operational definition as a method of assigning meaning to constructs; third, the use of conversational language.

The assumption of reality status allows one to postulate an unknown set (yet-to-be-discovered set) of properties and relations as possibly influencing any given set of obtained data. Further, there is no way to determine which constructs may be legitimately said to be operating, hence there is no way to make proper deduction of expected experimental results. In fact, it makes the prediction of phenomena on the basis of prior deduction from independent fact impossible. This, of course, leads to *ad hoc* explanation—which must, by the nature of the assumption of surplus meaning, be the only form of theorizing possible. The outcome of such "ad hoc" ism is apparent when one views the stagnation of psychoanalytic theory from the point of view of scientific endeavor. A theory cannot proceed from undefined gross analysis to a precise quantified science. For an opposing point of view see Marx (10). The nature of the formulation of such theory prevents development. It precludes the possibility that the theory will be able to predict erroneously—thus the theory is always supported, regardless of experimental outcome.

The use of discussion instead of the operational definition to get across what is meant by a given construct places undue emphasis on rational or intuitional interpretation of experimental results, and leaves any operational defini-

tion of the construct up to the experimentalist who wishes to work with it. Whether the construct is thus correctly formulated and interpreted in a given experimental situation is not determined by the theory of which the construct is a part. As a consequence, claims of misrepresentation, distortion of meaning, and destruction of purpose, are occasionally leveled at the experimentalist. This amounts to a state of authoritarianism in the less rigorously formulated areas of psychology. If meanings are assigned by operational definition, the theorist leaves only the task of validation to the experimentalist. The accuracy and success of a theoretical formulation are open to the public—indeed, even to the neophyte in psychology!—for confirmation or denial.

The third source of surplus meaning lies in the use of conversational language as a method of conveying meaning. The surplus meaning is admitted through the fact that the words (expressions) of a conversational language have meaning in context, meaning which varies as the words are used in different sentences. The use of constructs having operational meaning is simply a device for achieving fixed and precise meaning in a communicable and deductively manipulable form, and eliminating surplus meanings and their consequences.

In view of the above discussion we state the following rule of procedure: *To be testable a theory must be formulated as complete at the time of the test.*

5. *Rules for the admission of a construct.* Keeping the above discussion in mind, we should like to list what seem to us to be necessary and sufficient conditions for the admissability of constructs in theory construction. We do not wish to imply that any construct will necessarily be successful if it meets the criteria we shall set down. Success or lack of success is to be determined by empirical test of the content of the theory in question. The constructs in a theory must meet the following criteria if the theory is to be testable.

First, a construct which is adequate for use in empirical theory construction is operationally defined. The operational definition contains *a description of manipulations which are performed in the laboratory, and a description of the effects of these manipulations upon the behavior of the experimental subjects.*

Second, the concept under consideration must bear *explicitly stated relationships to other constructs in the theory of which it is a part.*

Third, a construct must *vary unidimensionally and continuously, and must affect at least one abstracted aspect of behavior, such that the behavior will vary unidimensionally and continuously.*[2]

[2] We define measurement as the assignment of numbers as names of members of a range of attributes which objects possess, such that the magnitude of the number reflects the magnitude of the attribute possessed by the object. It will become obvious that with any given measurement procedure we can work with one and only one range of attributes at a time. This is to say that one measurement procedure will define different numbers in one and only one class of qualities. Measurement consists in discovering which member of a class of qualities is possessed by a given object of measurement. Thus if the construct in question is to be testable, it must conform to the assumptions of the real number system. This is the reason we assert that a construct must vary unidimensionally. The construct will take different values in different situations, but all values will represent (name) qualities which are members of only one class of qualities.

In addition to the above there is the further consideration that where behavior is complex and discontinuous, there the theory will also be complex. If we observe a discontinuity in behavior in a given situation, it is reasonable to assume that the behavior is not a manifestation of a single construct.

As an elaboration of certain implications of the first and second criteria above, we list the following two statements: (*a*) To avoid circularity, a construct must be used only to predict behavior which is independent of its definition. (*b*) The construct is logically prior to the behavior which it predicts.

The notion of logical priority makes no assumption about temporal priority, i.e., antecedent events. The notion of temporal priority is not to be taken as a restriction upon empirical constructs. Its use as a restriction upon empirical constructs arises from thinking about theory from a particular point of view; namely, the metaphysical assumption of a causal relation between antecedent events and consequent events, and from the pragmatic utility of predicting future events. The only restrictions we wish to make are that predictions about behavior shall be deducible from the laws from the theory, and that the particular constructs which mediate the deduction be defined independently of the behavior which they predict.

These conditions for acceptable constructs are not new with the present authors. They are commonly found scattered throughout the philosophy of science, and in discussions of procedure in psychology.

In closing, we should like to remark that empirical science, indeed, science in the inclusive sense, is an activity of human beings. The constructs with which the empirical scientist works are the products of his own imagination, imagination which is guided by knowledge of empirical events. The criterion of a successful theory is its adequacy as an instrument for prediction and explanation. Nothing is inconceivable, if it allows us to predict successfully and accurately. The constructs with which the empirical scientist works have those properties, and only those properties, which the scientist gives them in any particular formulation of a theory. Failure to predict successfully is an indication that the theory is either inadequately formulated, or its empirical content needs to be changed in some particulars. But, to assume that constructs have surplus properties which are not accounted for is another way of saying, "Psychology is an art, not a science."

REFERENCES

1. BERGMAN, G. An empiricist's system of the sciences. *Sci. Mon.*, 1944, **59**, 140–148.
2. BORING, E. G. *A history of experimental psychology.* (2nd Ed.) New York: Appleton-Century-Crofts, 1950.
3. CARNAP, R. Testability and meaning. *Phil. Sci.*, 1936, **3**, 419–471.
4. CARNAP, R. Foundations of logic and mathematics. *Int. Encycl. unif. Sci.*, Vol. I, 1949, no. 3, 1–8.
5. EINSTEIN, A., & INFELD, L. *The evolution of physics.* New York: Simon and Schuster, 1938.
6. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
7. KRECH, D. Dynamic systems, psychological fields, and hypothetical constructs. *Psychol. Rev.*, 1950, **57**, 283–290.
8. LEWIS, C. I. *Analysis of knowledge and valuation.* LaSalle, Ill.: Open Court Co., 1946.
9. MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.
10. MARX, M. H. Intervening variable or hypothetical construct. *Psychol. Rev.*, 1951, **58**, 235–249.
11. MOWRER, O. H., & LAMOREAUX, R. R. Conditioning and conditionality (discrimination). *Psychol. Rev.*, 1951, **58**, 196–212.
12. WHITEHEAD, A. N. *Process and reality.* New York: Macmillan, 1929.
13. WOODGER, J. H. The technique of theory construction. *Int. Encycl. unif. Sci.*, Vol. II, 1947, no. 5, 1–13.

[MS. received May 21, 1952]

Psychological Review
Vol. 60, No. 3, 1953

THE BRAIN ANALOGY: ASSOCIATION TRACTS[1]

H. EDGAR COBURN

Registered Civil Engineer, San Diego, California

There are four phenomena that indicate a need for a more complex cortical structure than that shown in the original Brain Analogy paper (2, p. 156). These phenomena are secondary conditioning (4, p. 33), conditioned inhibition (2, p. 175), sensory preconditioning (1), and transfer of differentiation (4, p. 228). The present paper uses the first three phenomena as examples. Conditioned inhibition can be explained by the original structure, but since secondary conditioning cannot, and the two phenomena have almost identical experimental programs, both are described in terms of the new assumptions.

In addition to the above there are several other facts which indicate that the original Brain Analogy (BA) hypothesis is not without its difficulties. Pavlov has noted that repeated application of tone *A*, say, does not result in differentiation unless the method of contrast is used—tone *B* without reinforcement (4, p. 117). The original hypothesis (2, p. 173) easily meets this test in the laboratory. But transfer the BA to a natural environment between experiments and differentiation takes place without the formal method of contrast because other specific neurons and the generalized neurons are actuated by many natural auditory stimuli which, of course, are not reinforced by the appropriate US. In time, even the specific conditioning to tone *A* would be lost.

An interesting fact deduced from the above situation is that monopolar neurons are relatively useless for complex animals. Since it is a fact that laboratory conditioning remains fairly stable in magnitude when exposed between experiments to natural stimuli, most neurons are probably bipolar with one pole representing some uncontrolled factor in the laboratory. The absence of this factor between experiments prevents extinction.

Generalized neurons are inevitably involved in any problem concerning differentiation and their properties are necessarily related to the complexity of the environment. A simple mechanism suffices for a simple environment but added factors require a more complex structure for both generalized and specific neurons.

These facts indicate that the difficulties in the original hypothesis were of degree rather than principle—the machine's IQ was too low to cope with two environments. Since intelligence increases with the resolving power of the analyzers (3, p. 456), intelligence is low where monopolars predominate. However, this paper is based on a laboratory environment which adequately illustrates certain principles.

The original BA structure allowed conditioning from sensory neurons to motor neurons. The new BA structure has the additional property of permitting conditioning between sensory neurons. A sensory neuron is now defined as *a unit consisting of one or more receptor elements or poles, a latency element in each pole, a combining element, a cell body, and a multitude of branches each terminated by a delta cell which connects with a motor cell body or with the cell body of another sensory cell.* A generalized neuron is defined as *a unit*

[1] The author wishes to express his gratitude to Vera Jane Coburn for making the drawings and for assistance in preparing the manuscript.

consisting of a multitude of receptor elements having identical properties, including latency, a combining element, and a multitude of branches each terminated by a delta cell which connects with the cell body of a sensory neuron. The connections to sensory cell bodies are known as association tracts (AT's). Because the mechanism of generalization is restricted to the AT's, it can affect the motor cells only in an indirect manner through conditioning to sensory cells (AT conditioning) which in turn connect with the motor cells (sensory-motor, or SM, conditioning). The purpose of this provision is to permit transfer of differentiation with little or no interference by generalization.

The AT's make it possible for two signals to become associated more or less independently of the activity of the motor cells—independently of conventional reinforcement.

In the original paper conditioned inhibition was attributed to the so-called zero-ratio bipolar neuron (2, p. 175). This naïve notion, suitable for a first approximation, is quite limited in application. We now observe that since thresholds have normal distribution (2, p. 157), any bipolar neuron which has an inhibitory pole stimulated at a subthreshold level, in effect complies with the zero-ratio postulate. But there is an additional advantage in that the available population of *nonactuated* bipolar neurons is roughly proportional to the stimulus strength at the "zero" pole. This is the mechanism which allows conditioned inhibition to function best when the conditioned inhibitor is strong (4, p. 74) because a neuron that is not actuated on a nonreinforced trial possesses stable conditioning and the strong stimulus insures a sufficient population of these. A weak stimulus causes an insufficient population of stable bipolar neurons, allowing competition by gamma phase differentiation[2] (2, p. 170) of monopolar neurons to dominate the response. Secondary conditioning then reveals itself through means of AT connections.

Although the new structure is more complex, the principles are familiar to readers of the first paper (2) and the burden of additional complexity is more than offset by the increased scope and power of the technique. While this paper is limited to only three examples, one can easily visualize probable applications to such theoretically important subjects as latent learning, token rewards, and others.

THE STRUCTURE

The analysis is somewhat simplified by the fact that the experimental program automatically collects certain groups of neurons into functional clusters and eliminates others from consideration. But this is offset by the existence of several hundred groups. Since an exhaustive analysis is possibly unnecessary at the present time, precision in proof is sacrificed for convenience in exposition.

Every problem involving two overt stimuli can be placed in one of two categories: The overt stimuli actuate certain *receptors* in common; or, they do not. If overt *A* and *B* are dissimilar enough, auditory and visual for example, to avoid actuation of any common receptors as shown in Fig. 1, the number of groups is a minimum. In this paper, generalization is assumed insufficient to permit either overt stimulus to act as a substitute for the other.

Neurons with two excitatory poles must have both actuated to get a response from the combining element, but the neurons with an inhibitory pole will respond only when the threshold of the

[2] Formerly called *periodic reconditioning* (2, p. 162).

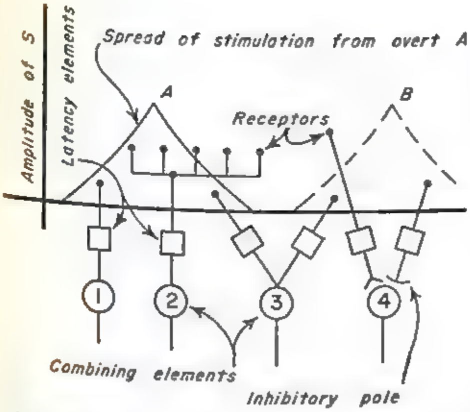


FIG. 1. The location of the receptors in the figure is significant as to threshold and position in the sensory field. The receptor of the excitatory pole of neuron *4*, for example, has a relatively high threshold while the receptors of *3* have low thresholds. The excitation of any receptor is usually a function of stimulus position as well as intensity.

inhibitory pole is not exceeded. However, in every bipolar case due consideration must be given the latency elements. Simultaneous stimulation of both poles is not an adequate criterion: the effect of the stimulus must be simultaneously present at the combining element.

Each group, represented by a single neuron in the figures, consists of many individual neurons with the usual distribution of properties (2, *Postulates*) except, of course, that if a group has specified threshold limits, it represents only a particular fractional part of the area under a normal distribution curve. It is essential to understand that any alteration in an experimental program, such as a change in magnitude of *A* or *B*, reclassifies many of the groups and transfers some to other functional clusters.

Certain neurons, sometimes structurally dissimilar, are functionally identical in a particular experimental program. Neurons *1* and *2* of Fig. 1 are functionally identical, except possibly for latencies, because they are both actuated whenever *A* is present and neither is ever actuated by any other phase of the indicated stimulus situation. In general, then, we can classify the neuron groups according to their behavior in the experimental program without concern for their structures. These functional clusters are relatively few in number.

There is, however, one outstanding peculiarity of the new mechanism: The AT conditioning has little effect on the motor response unless the program is altered.[3] The experimental alteration, of course, is the test trial which reveals the existence of AT conditioning. The cause of this indifference to the presence of the AT's is found in the program itself. Any AT conditioning to sensory cells which are *not* reinforced by the US is "open circuited" at the motor cell body whereas any AT conditioning to sensory cells which *are* reinforced is superfluous. But let a change be introduced into the program and the presence of AT conditioning is immediately apparent. For example, as long as AT activity is accompanied by SM activity, there is little or no experimental evidence that the CR is partly due to the former, but when the latter is absent, as on a test trial, the CR is still present and attributed by the hypothesis to AT conditioning. The interference of pulses in the sensory cell body and the reduction in the *E* function (2, p. 171) due to extra stimulation, combine to maintain an output pulse rate that is considerably less than the sum of the inputs. It is concluded that with a constant program the AT's are without significant effect on observed behavior. Figure 4 will aid in analyzing the above argument.

[3] The conditioning and extinction properties for AT connections are informally assumed to be the same as for SM connections. The latter are given in detail in the original BA paper (2).

It is also seen from examination of Fig. 4 that if the program is arranged so that AT conditioning occurs from each of two sensory neurons to the other, and this is possible, a regenerative or positive feedback circuit is established which might continue to operate once started. No assumptions have been made concerning latency in the AT's, but it is apparent that a finite value exists which could cause a train of pulses to appear in the circuit. However, a short conduction time causes the initial pulse to be returned to the starting neuron while it is still refractory following the generation of the pulse (2, *Postulate 10*) and immediately stops the feedback action. Since the delta cell is not reinforced, extinction ultimately takes place, breaking the feedback circuit.

SECONDARY CONDITIONING

Secondary conditioning has been defined by Pavlov (4, p. 33) as the CR which develops when a neutral stimulus is followed by a CS that is not reinforced on the occasional secondary-conditioning trials. After practice the neutral stimulus, alone, evokes the UR. It is an essential condition that the secondary CS be weak in fact or in effect (by delay in applying primary CS). The experimental evidence is based on dissimilar overt stimuli as in Fig. 1. Tone *A* is the primary CS while *B* (visual) is the secondary CS. The US is never paired with *B* which is *followed* by *A*, alone.

In the experiments reported by Pavlov the primary conditioning was established prior to any secondary-conditioning trials. Brogden's work on sensory preconditioning (1) shows that if the secondary conditioning is established first, the secondary CS does not have to be relatively weak and terminated prior to the start of the primary CS. The BA analysis reveals that the necessity for the restrictions in the Pavlovian example is the interference by the phenomenon known as conditioned inhibition. Such interference is impossible in Brogden's experiments. These properties of the cortical mechanism will be discussed more fully in later sections of this paper.

Figure 2 shows that the secondary CS, though applied alone in its overt form, overlaps the action of the primary CS at the delta cells owing to the distribution of latencies. It also indicates that the effective duration (2, p. 159 and 3, p. 457) of *A* is prolonged for the same reason. Although the latency distribution function is continuous (2, *Postulate 5*), we have simplified the analysis by assuming that latencies take a discrete set of values, each one of which includes a number of neurons. Figure 2 is based on the latency ogive (2, Fig. 6) with the ordinates divided into steps of 2 per cent each, but with a linear time scale.

Since the short-latency neurons of the secondary CS (B_1 to B_{44} incl.) are never actuated simultaneously with the primary CS, they can never become conditioned and are henceforth excluded from the discussion. The long-latency secondary CS (B_{48} to B_∞ incl.) are excluded because they are actuated subsequently to those primary CS (A_1 to A_{41} incl.) which are actuated early enough in the program to become conditioned to the US: conditioning may occur between long-latency secondaries and long-latency primaries (A_{42} to A_∞ incl.), but this fact is immaterial because the motor cells cannot have stable connections except with delta cells which have their activity initiated prior to motor activity (2, p. 169). Backward conditioning is excluded. Our concern, then, is only with medium-latency secondaries (B_{45} to B_{47} incl.), short-latency primaries (A_1 to A_{41} incl.), and forward conditioning.

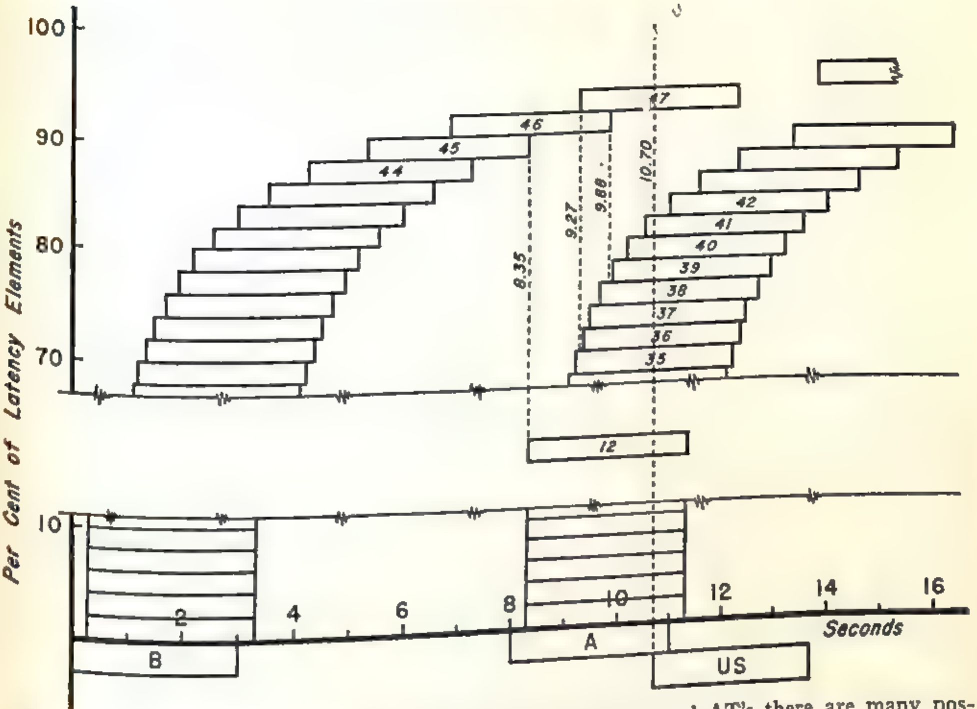


FIG. 2. Because sensory neurons have both latency elements and AT's there are many possible combinations of channels for signals in their traverse of the cortical mechanism. On the other hand, since conditioning is possible only when the appropriate contiguous elements are simultaneously in action, many other channels are closed by any particular experimental program. This figure is the Brain Analogy version of the long-established "persistent trace" concept applied to two stimuli to show how the long latency neurons of one stimulus are simultaneously in action with short latency neurons of another stimulus.

The latencies shown in Fig. 2 apply to the poles of the four types of neurons considered in this paper: Monopolar, generalized, excitatory bipolar, and inhibitory bipolar.

Because the latencies overlap in various degrees, certain groupings are possible on a basis of latency. Aggregate B_{45} can be a CS for A_1 to A_{11} inclusive, and to B_{46}, while B_{46} can be a CS for A_1 to A_{38} inclusive, and to B_{47}, which in turn is a CS for A_{36} to A_{41} inclusive. On the other hand, A_1 to A_{35} inclusive can also be CS for B_{47}. These relationships assume monopolar neurons. Bipolars with two excitatory poles will be actuated for no longer than the duration of the shorter stimulus and usually for less time because of failure of the timing to be optimum. Bipolars with one inhibitory pole will usually be excited part of the time unless the stimulus for the inhibitory pole is of greater duration or precisely timed. Of course if the inhibitory stimulus has less duration than the positive stimulus, no inhibitory bipolar can be suppressed for the entire cycle.

The analysis now proceeds on the basis of a restricted sample of the latencies shown in Fig. 2. Suppose we consider only B_{47}, A_{36}, and A_{41}. The temporal relationship of monopolars, bipolars, and reinforcement is shown in Fig. 3. The cross-hatched areas are of particular significance: these are the inhibitory bipolars that are not actuated on nonreinforced trials and hence possess stability. It will be observed, however, that the inhibitor, B_{47}, does not

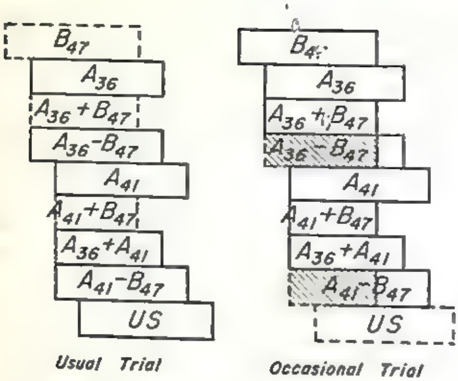


FIG. 3. Latencies apply equally well to monopolar neurons and to the individual poles of bipolar neurons. The sample mentioned in the text is shown here to chronological scale. In this figure only, the algebraic signs, plus and minus, are used to indicate excitatory and inhibitory poles, respectively. The usual trial is reinforced by the US to establish the primary CS. Occasional trials are nonreinforced (by the US) to establish, by differential action, the secondary CS or conditioned inhibitor. The cross-hatched areas represent the nonactuated portion of a trial where failure of reinforcement is without effect; these are relatively stable. But for successful conditioned inhibition the duration of B must exceed A to prevent "unmasking" of the inhibitory bipolar and resultant secondary CR.

suppress all of the action of A_{36} and A_{41} because the former lacks effective timing or sufficient excess duration to compensate. The equal duration of overt A and B was intentionally chosen to illustrate this effect. The inhibitory bipolars, then, are only relatively stable as a group; but the possession of a sufficiently long gamma phase renders them unconditionally stable in an appropriate program and the presence of inhibition for part of the cycle makes the gamma-phase requirement easy to meet.

Omitted from Fig. 3 are three generalized neurons, one of each latency, which receive stimulation identical to that given the corresponding monopolars —physical separation of A and B (see Fig. 1) prevents differential competition between generalized neurons and monopolars for AT conditioning. Also omitted are A_{36} *without* A_{41}, an inhibitory bipolar, and its converse because though functional for part of the cycle, they are inhibited during the effective duration of the US and therefore cannot serve as an outlet for B. Since the reader might question why they are postulated at all, an explanation is in order. While the overt stimuli are quite arbitrary, the receptors are distributed in various tissues without concern for any artificial boundaries and consequently the labels on these neurons are experimental artifacts due to chance positioning and intensity of stimuli relative to the location of the receptors. Because of this fact the "one field" inhibitory bipolar is accounted for if only to exclude it safely.

Technically, B_{47} *without* A_{41}, an inhibitory bipolar, because it can be a CS to *1* and *6* of Fig. 4 should be included. However, since it is never reinforced by the US and is essentially the same as B_{47}, it is omitted to simplify the figures. The case of B_{47} *without* A_{36} is different: it is always inhibited by the time any other neuron in the sample is excited and therefore properly excluded.

And, finally, we have the *A or B* monopolar category. In this program the two cases are excluded because they represent an impossible condition with fields of stimulation that do not overlap (Fig. 1).

Starting with 17 cases, we have eliminated 5 for necessity and 4 for convenience. The remaining 8 with the US are shown in Fig. 3. However, in Fig. 4 we introduce one of the generalized neurons to facilitate understanding. The conditioning symbols are the same as used previously (2, p. 174) and are not repeated here since the present paper is unintelligible without the first. It is important to understand that the individual neurons in Fig. 4 represent functional clusters rather than structural counterparts. An inhibitory bipolar, for

example, is functionally a monopolar if the inhibitory threshold is not exceeded. An experimental program may reduce a complex neuron to one of simpler function but, of course, the reverse is never true.

It will be seen in Fig. 4 that the experimental program develops many AT conditioned connections as well as the SM connections. As noted earlier, the AT conditioning has little effect on the motor response unless the program is altered. The experimental alteration, of course, is the test trial—application of *B*, alone—which is used to reveal secondary conditioning. Since the connections from *4* through *1* to *2* and *5* must be both functional and actuated to have any effect on the motor cell, their existence is hidden by the presence of *A* on every trial except the test trial. In addition, neuron *4* is never actuated, except for the test trial, under conditions which cause it to influence, independently, conditioning between *1* and the motor neuron because the program forbids the required sequence of events. Another point of sig-

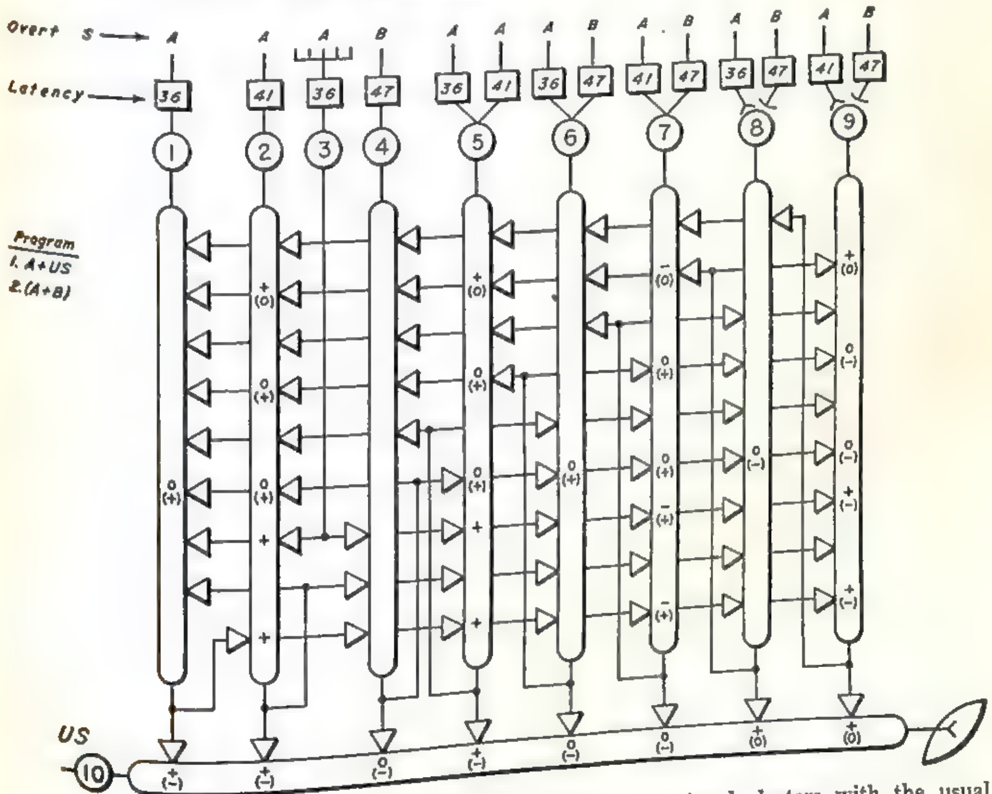


FIG. 4. The individual neurons in this figure represent functional clusters with the usual threshold variations. As *B* is varied in strength there is a shift in relative populations. This variance, in conjunction with the existence of the AT's and other factors, is the responsible agent in determining whether secondary conditioning or conditioned inhibition will be observed. It should be noted that the AT delta cells transmitting the output of any sensory cell are actually in parallel, not apparently in series as shown. The conditioning symbols used here really apply to a case where the duration of *B* considerably exceeds *A*. The significant point is that no matter what the temporal relationship of overt *A* and *B*, provided *B* starts first and has sufficient duration, there are always combinations of latencies for inhibitory bipolars which keep some of them totally inactive whenever the conditioned inhibitor is presented as part of the program.

nificance is the fact that no conditioning occurs from *4* to *8* or *9*.

The circumstances of the program are such that neurons *1*, *2*, and *5* are in competition with *8* and *9* for control of the motor neuron. If *B* is weak, the population of *8* and *9* will be small—not structurally, but functionally. The inhibitory bipolars that have inhibitory thresholds in excess of the overt *B* are functionally identical with monopolars excited by *A*, alone. It is seen then that the populations of *1*, *2*, *5*, *8*, and *9* vary relatively with stimulus intensity; this variance is the responsible agent in determining whether secondary conditioning or conditioned inhibition will be observed (4, p. 69). The fact that *8* and *9* also have access to the motor cell when *B* is weak is of no qualitative importance, *unless 1, 2, and 5 are excluded by differential action*, because *B*, the test for secondary conditioning, and *A plus B*, the test for conditioned inhibition, both have outlets through *1*, *2*, and *5* which prevent the appearance of overt conditioned inhibition. This is possible because conditioned inhibition is not an active suppression of any CR. The result is secondary conditioning since *4*, on a test trial, evokes *1* which evokes *2* and *5* through AT connections and all three unite in generating the second-order CR.

Certain aspects of the explanation possibly need amplification. The first part of the program consists of repeated trials of *A* followed by the US. The neurons represented by *2* are the objects of competition by AT conditioning from *1*, *3*, and *8*, which in effect, are components of a compound stimulus as viewed by *2*. We have no data upon which to assign relative strengths to the components except to observe that since overt *B* has been assumed to be weak, *8* must be a very small component because *8* represents only a small fraction of the sum of the populations of *8* and *9*: the sample fails to indicate that both *1* and *8* represent a few of the shorter latencies of their respective categories. Most of *2*, therefore, are appropriated by *3* and actuation of the latter reveals generalization. The connection from *8* is insignificant and the tract from *1* is of intermediate importance. The last part of the program consists of the aforementioned trials interspersed with occasional trials of overt *B* followed by *A*. Because nearly all of the *2* group have already been appropriated, AT connections from *4* are disabled and denied access to *2*. But access to *1* is available; this connection, on a test trial, produces the second-order response through *2* and the motor cell. The behavior of *5* parallels *2*.

By delaying *A*, only the tail-end of the latency distribution of *B* is available for conditioning and hence the effect on relative populations is the same as though overt *B* had been reduced in magnitude because the number of neurons actuated by *A* remains unchanged. This is more evident when it is remembered that the sample being considered is one defined by coexistence with the effective duration of the US.

In conformity with the evidence, secondary conditioning is observed only when *B* is weak in fact or effect.

CONDITIONED INHIBITION

If *B* is a strong stimulus, the population of *8* and *9* will be relatively large because more bipolars will have their inhibitory poles actuated (more thresholds will be exceeded) and therefore a greater population will fail to respond on nonreinforced trials. These have stability. On the other hand, *1*, *2*, and *5*, sometimes actuated without reinforcement, lack stability, and being of relatively small population cannot depend on gamma-phase differentiation which incurs an additional loss of conditioned

delta cells. They surrender control to *8* and *9*.

Overt B now has no outlet through *1*, *2*, or *5* and its application, alone, excites no overt response. When A is applied, alone, the response is due to *8* and *9*, not as originally established through *1*, *2*, and *5*. In addition to the direct action of *8* and *9* through its own receptors, there is a possible action through AT conditioning from *1* and *3* to *9*, but these unstable connections are gradually displaced by AT conditioning from *8*. When A and B are actuated in conjunction, there is no overt response. This last is called *conditioned inhibition* though it has not been shown to be an active suppression of any CR. On the contrary, *the original CR is driven to extinction, rather than suppressed, by a competitor possessing stability and in a process identical with other forms of differentiation.*

It is apparent that conditioned inhibition is favored not only by a strong conditioned inhibitor, in conformity with the evidence, but also by one of considerable relative duration. The latter fact appears to have escaped attention. It is equally apparent that AT conditioning plays a significant part in the isolated phenomenon of conditioned inhibition by causing the block of inhibitory bipolars to act as a functional unit—the AT conditioning from the shorter-latency bipolars displaces AT conditioning from other neurons, especially the generalized group.

If the animal were prepared surgically or by suitable nerve blocks in such a way that influence by the musculature system (kinesthetic receptors, etc.) could not influence the experiments, it seems probable that a smooth transition would occur between conditioned inhibition and secondary conditioning as B was varied in separate experiments from one intensity extreme to another. We are concerned, of course, with the salivary reflex. Pavlov has reported the existence of "defense" reactions at the transition intensities (4, p. 69); these, by alteration in the Q and E functions (2, *Postulates 12* and *15*), possibly reduce the amount of conditioning in both groups to an extent that makes observation difficult.

SENSORY PRECONDITIONING

Sensory preconditioning has been defined by the experiments of Brogden (1, p. 325) as the conditioning which forms between two stimuli actuated simultaneously without conventional reinforcement. The existence of the bond is revealed later when one of the stimuli is used as a CS with some US to develop a CR, and a test trial of the other stimulus then also shows a connection with the motor cell by exciting the UR with which it had never been associated.

Following the general scheme of the preceding sections we assume A is the auditory stimulus and B is the visual one. Each trial consists of the simultaneous actuation of A and B for three seconds. After a number of such trials A, alone, is followed by shock, the US, for enough trials to establish conditioned leg flexion. A test trial with B, alone, now produces leg flexion—a response with which B never had been associated.

Our sample this time consists of, say, A_{40} and B_1. The subscripts represent the 2 per cent ordinates, as before, except that the starting time of A and B is simultaneous. In the first set of trials AT conditioning forms from B_1 to A_{40}. In the second set of trials SM conditioning forms from A_{40} to the US. A test of B_1 then evokes A_{40} which in turn evokes the UR.

It will be noted that this hypothesis is not limited to the simultaneous action of overt A and B. They can be successive as in Fig. 2. Suppose B_{47} conditions to subsequent A_{40} in the AT's

as before. Then if overt *A* is conditioned to the US, application of overt *B* will evoke the UR through AT conditioning. But if the role of *A* and *B* is reversed, the temporal relationship becomes more critical though not qualitatively different. Again following Fig. 2 we see that a short-latency *A* will condition to a subsequent long-latency *B*. Then if overt *B* is conditioned to the US, application of overt *A* will evoke the UR through the AT's, but only if the program had been arranged so that the US followed the beginning of the long-latency *B;* otherwise, the motor cell would be appropriated by a short-latency *B* to which *A* is not a CS.

Brogden's experiments were concerned with a musculature reflex and with avoidance learning. In the early stages of such conditioning many CR's form (3, p. 458) and it is only with practice that a particular CR heads the hierarchy. When the other CR's become extinct, they "open-circuit" the corresponding AT conditioning so that the differentiation is evident from the first test trial.

It was noted in connection with secondary conditioning that while inhibitory bipolars of sufficient population convert the phenomenon into conditioned inhibition, such interference is impossible in Brogden's experiments. This is true because the inhibitory bipolars corresponding to *A without B* are not actuated in the first set of trials which establish AT conditioning; in the last set of trials they compete with *A* for control of the UR but not in any differential manner that could cause one or the other to show instability. However, the fact that these inhibitory bipolars appropriate some of the motor cells is not without effect: The outlet that *B* would otherwise have through AT conditioning to *A* and US is correspondingly reduced. This agrees with Brogden's observations but, of course, we have not shown that this is the sole cause for the reduced response from *B*.

Conclusion

The AT postulates appear to offer significant advantages in correlating behavior phenomena; however, they present so many unexplored and inadequately explored facets that we can hardly expect freedom from error. But if they are fundamentally sound in *principle*, it seems probable that they can be improved in detail. Possessing a physical basis like the other Brain Analogy concepts, they provide all the research advantages of a *tangible* structure.

REFERENCES

1. Brogden, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, **25**, 323–332.
2. Coburn, H. E. The brain analogy. *Psychol. Rev.*, 1951, **58**, 155–178.
3. Coburn, H. E. The brain analogy: a discussion. *Psychol. Rev.*, 1952, **59**, 453–460.
4. Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep.) London: Oxford Univer. Press, 1927.

[MS. received May 27, 1952]

Psychological Review
Vol. 60, No. 3, 1953

ONE- AND TWO-TAILED TESTS

MELVIN R. MARKS

AGO, Personnel Research Section

Hick, in "A Note on One-Tailed and Two-Tailed Tests,"[1] purports to vitiate the recommendations made by me in a recent paper.[2] I think a reply is indicated.

1. My paper was not intended to precipitate a controversy over the philosophical foundations of mathematical statistics. It *was* intended to show that, by taking advantage of available statistical techniques, the experimenter could increase the precision of his investigation by recognizing and minimizing the incidence of Type I and Type II errors. The paper was expository rather than argumentative, a restatement rather than a presentation *de novo*. All statements made therein relative to the testing of hypotheses have the acceptance of statisticians generally.

2. The experimenter's decision to adopt a particular level of confidence is not made with reference to statistical considerations. Such a decision depends entirely on the assessment of the likelihood of Type I and Type II errors and the practical importance to be attached to their occurrence. Thus, the propriety of the one- or two-tailed test depends only upon the nature of the hypothesis to be tested—not on the level of confidence adopted.

3. Although logic is timeless, the "personal equation" of the experimenter is not. The legitimate use of particular data for the test of a particular hypothesis *does* depend on why (the basis for), if not when, the hypothesis was formulated. Hick says, ". . . it does not and cannot matter whether a theory (read hypothesis) was conceived before, during, or after the experiment; it may be suggested by the data, or it may be revealed in a dream." Now this statement is true in *all* respects *if* conclusions about the tenability of the hypothesis are to be restricted to the data at hand. But if these conclusions are to be extrapolated, we must strike the phrase, "it may be suggested by the data." For consider, with reference to any particular data, if the hypothesis is to proceed from the data, the best hypothesis is that the data are as they are. In such case, we might solemnly aver that they are as they are by definition, and never bother to test any hypothesis. We shall always be right, if only we select the hypothesis carefully enough—until the next time!

4. Hick challenges, at least implicitly, the Neyman and Pearson theory as it applies to critical regions. The pros and cons of such a controversy have no place in this discussion. However, it may be stated that, when the critical region selected is appropriate to the hypothesis being tested, then the level of confidence (by definition) is known exactly, and the probability that a Type II error will occur (false acceptance of the null hypothesis) is minimized.

5. Hick's remark on my treatment of the chi-square test has merit. Technically, the term "two-tailed," as applied to the chi-square distribution, refers to the actual tails of that distribution—i.e., the regions of exceptionally large and exceptionally small values of chi square. However, it is still legitimate to cut the level of confidence in half

[1] W. E. Hick, A note on one-tailed and two-tailed tests. *Psychol. Rev.*, 1952, 59, 316–318.

[2] M. R. Marks, Two kinds of experiment distinguished in terms of statistical operations. *Psychol. Rev.*, 1951, 58, 179–184.

when we predict direction of frequency discordance. This is so because when we eliminate exactly half of the possible values of chi square (that half corresponding to frequency discordances in the unwanted direction), we also eliminate half of the Type I errors.

6. The same remarks apply, with somewhat less force, to the indicated use of the F test. Bendig[3] has pointed out that only one tail of the F distribution is tabled—i.e., values of F equal to or greater than unity. When I discussed the increased precision which could be achieved by predicting the hierarchy of magnitude of the means, I was referring to the fact that, in the case of simple classification analysis of variance with only two columns (when $t^2 = F$), the number of Type I errors is cut in half when the negative values of t are eliminated beforehand. The number of Type I errors would be reduced still further in the case of more than two columns (variables) if the ordering of the means was predicted beforehand, although in such case power would suffer—i.e., the chances of committing a Type II error would increase since a slight inversion from the predicted order would necessitate rejection of the experimental hypothesis.

[3] Personal communication.

[MS. received August 8, 1952]

Psychological Review
Vol. 60, No. 3, 1953

IDIODYNAMICS AND TRADITION

SAUL ROSENZWEIG

Washington University
St. Louis, Mo.

By an *actual* historical accident the recent paper by Seeman and Galanter (4), attacking the writer's "Idiodynamics in Personality Theory" (2), appeared in the same issue with his "The Investigation of Repression as an Instance of Experimental Idiodynamics" (3). The latter article goes a long way toward a reply; only two or three additional comments are required.

1. Seeman and Galanter mistakenly assume that idiodynamics was considered to be wholly embodied in the projective methods; or, again, that the projective methods are a necessary part of the idiodynamic approach. The historical accidents which they cite are, indeed, accidental to the argument since idiodynamics, while implicit in and exemplified by the projective methods, extends far beyond them. The relevance of *experimental psychology* has been shown in the above-cited reference (2); the kinship to *psychotherapy* is found in the phenomenological approach now widely current. Under these circumstances the arguments presented by Seeman and Galanter which demonstrate that idiodynamics is not confined to the projective techniques are gratuitous.

If, as they also aver, the MMPI and other structured instruments are being developed along configurational lines, and the psychology of learning is, with the help of Skinner, recognizing the importance of emitted responses (response dominance), the conclusion to be drawn is that the orientation called *idiodynamic* is by no means limited to one aspect of psychology. It was for this very reason of breadth that the idiodynamic position was outlined in the original paper as a theoretical basis for the empirically overdeveloped projective methods; but at no point was it there maintained that these techniques exhausted the theory or that it arose from them.

2. The critics' appeal to history demands scrutiny. It is to Skinner, they insist, that the revolution in respect to response dominance must be attributed. They write (4, p. 289): "We have shown that with respect to the 'postulate' of response dominance, the 'revolution in the conception of the stimulus' was instigated by Skinner rather than by psychodynamic or 'idiodynamic' psychologists." Here they are factually in error and the evidence is indicated in the paper they criticize. It was there shown that Thurstone in *The Nature of Intelligence* (6, Ch. I) and earlier (1923), in this very journal, cogently expounded his "stimulus-response fallacy." In these terms he takes to task the assumption that all behavior can be accounted for and controlled by the stimulus alone and forcibly calls attention to the important and academically neglected relationships between drive and response *within* the organism. In partial support of his position he quotes from Jennings as follows:

> Activity does not require present external stimulation. A first and essential point for the understanding of behavior is that activity occurs in organisms without present specific external stimulation. The normal condition of Paramecium is an active one, with its cilia in rapid motion; it is only under special conditions that it can be brought partly to rest. Vorticella, as Hodge and Aikins showed, is at all times active, never resting. The same is true of

most other infusoria, and, in perhaps a less marked degree, of many other organisms. Even if external movements are suspended at times, internal activities continue. The *organism is activity*, and its activities may be spontaneous, so far as present external stimuli are concerned (1, p. 22).

But more saliently Thurstone makes explicit acknowledgment of his indebtedness to the "new psychology" of that time—psychoanalysis. Commenting, in summary, on this source of his ideas and on the vaunted objectivity of academic psychology, he writes (6, p. 164):

It is my belief that the attitude, which is implied in the so-called new psychology, has given, in a relatively short space of time, more insight into human conduct than the thoroughly objective point of view which has in recent years become established in scientific psychology, and which has been borrowed from related objective sciences.

Fourteen years later Skinner (5) presented the same fundamental idea as "the emitted response," revising the concept of the reflex and offering his contribution as an elaboration of behaviorism. (Incidentally, from the present ground of history it is a remarkable fact—perhaps a coincidence—that while Jennings' book, used by Thurstone, was entitled *Behavior of the Lower Organisms*, Skinner's later volume was called *Behavior of Organisms*!) Since Thurstone moved from the stimulus-response fallacy to factor analysis of the inner man while Skinner immediately exemplified his "operant behavior" in the field of animal learning, the thinking of the latter became more familiar than that of the former to some psychologists. But as *history* in modern psychology goes, not Skinner but Thurstone (and before him Jennings and Freud) must be credited with the formulation of what has been newly christened *response dominance*.

3. Seeman and Galanter make much of the threat which idiodynamics was said to present to traditional psychology, and their quotations are culled largely to support this interpretation. (This note has even crept projectively into their "fictitious structured personality test" [4, p. 287] in which the two "personologically significant" items of the three cited by them from a limitless, random universe involve rejection and distrust. Here the reaction of the authors will, no doubt, be: "Most notags karolize elatically"—which is their remaining item.) They then proceed to take up the cudgels for "tradition." But the idiodynamic orientation, which they characterize as belonging to clinical psychology as a whole, is by no means negatively motivated and is not inimical to all tradition. In fact, the just preceding references to Thurstone and Skinner—to say nothing of Freud—make it clear that some of psychology's traditions are being articulated in, not threatened by, idiodynamics. Psychology is and has been many things to many men and it can well afford to remain so. For where if not among psychologists should the idiodynamics of thought be exemplified!

REFERENCES

1. JENNINGS, H. S. *Behavior of the lower organisms.* New York: Macmillan, 1906.
2. ROSENZWEIG, S. Idiodynamics in personality theory with special reference to projective methods. *Psychol. Rev.*, 1951, 58, 213–223.
3. ROSENZWEIG, S. The investigation of repression as an instance of experimental idiodynamics. *Psychol. Rev.*, 1952, 59, 339–345.
4. SEEMAN, W., & GALANTER, E. Objectivity in systematic and "idiodynamic" psychology. *Psychol. Rev.*, 1952, 59, 285–289.
5. SKINNER, B. F. *Behavior of organisms.* New York: Appleton-Century, 1938.
6. THURSTONE, L. L. *The nature of intelligence.* New York: Harcourt, Brace, 1924.

[MS. received August 19, 1952]

Psychological Review
Vol. 60, No. 3, 1953

KENDON SMITH'S COMMENTS ON "A NEW INTERPRETATION OF FIGURAL AFTER-EFFECTS"

CHARLES E. OSGOOD

University of Illinois

Comments, replies to comments, and often comments on replies to comments provide invigorating but seldom edifying contributions to our journals. They also take up space badly needed for new contributions. But since our first view of Dr. Smith's comments was on the pages of this journal (5), and since their tenor often implied misstatements on our part, I see no alternative but to make a public reply.

Smith justly criticizes us for failing to mention his critical note in the *American Journal of Psychology* on the satiation theory of the figural after-effect (4). Our only defense is that we were not attempting a complete review of the literature and did not come across his note in the other sources used. But let us look at the six points he raised there and repeated more recently:

1. *Neither satiation nor statistical theories explain after-effects like those obtained in the waterfall and Plateau spiral illusions.* These theories also do not explain contrast phenomena or the effect of values and motives upon perceived size—they are not required to cover all phenomena in the field of perception. Both of the illusions referred to involve continuous *movement* of contours in the field and may well involve mechanisms beyond area 17.

2. *Subjects who have worn distorting glasses for a long period show figural after-effects* (Gibson, 1). As I read Gibson (pp. 1–3), the effect of these prisms was such as to render vertical lines in the field curved (e.g., *I*-contours), and upon removal of the prisms truly vertical lines (e.g., *T*-contours) now appeared bent in the opposite direction. Both Köhler and Wallach (2) and Heyer and myself (3), each in our own way, deal with this situation explicitly, specifying how displacements arise. I fail to see the point of the comment.

3. *Gibson* (1) *obtained figural after-effects "when his subjects merely viewed curved-line inspection-figures."* And Smith adds that "there was little if any opportunity for configurational adaptation" in this case. I assume that what Köhler and Wallach called "self-satiation" is referred to here, e.g., that certain figures may show form distortion during original inspection. The latter authors do go to considerable pains to interpret this phenomenon, pointing out that the effect would be expected for curved-line figures, where resistance is denser on one side than the other, but not for straight lines.

4. *Similar effects have been observed in other modalities.* This is exactly what one should expect if either satiation or statistical models apply to sensory projection systems in general.

5. *After-effects occur across the vertical meridian of the eye.* Heyer and myself went to considerable lengths to spell out this schism between functional and anatomical data.

6. *After-effects "in the third dimension" have been reported.* Rather than merely resolving this issue "in an expression of scepticism" as Smith says, we (*a*) explicitly discussed how "depth" and "size" interpretations can be shown to be interchangeable in many studies where cues are ambiguous and (*b*) described a duplication of the most criti-

cal demonstration of this effect with negative results.

Specifically against the statistical theory Smith offers two criticisms. (*a*) He says that "in Köhler and Wallach's basic demonstration, where the inspection-contours *are* congruent with the test contours, displacement of the latter is commonly reported" (5, p. 402) and he cites p. 271. This would be contrary to the satiation theory as well as the statistical theory. What do Köhler and Wallach (2) actually say on this matter? "The figure which coincides with the previously inspected object will be pale or gray in comparison with its black partner, it will seem to lie further back in space, and *it may look a trifle smaller*" (p. 271, italics mine). That *here* this is an "interpretation" based on the paleness cues, that *displacement* really does not occur under conditions of coincidence of *I* and *T* contours, is clearly indicated by Köhler and Wallach at many points. Witness the following: "But as we bring the *T*-line nearer and nearer, the satiated *I*-region will gradually extend *beyond* the place of the *T*-line. . . . This development will continue until the resistance on one side of the *T*-line is just as great as it is on the other side. *This happens in the position of coincidence where no displacement can occur*" (2, p. 338, italics mine). At no point do Köhler and Wallach report evidence contrary to this statement; Smith may have failed to take into account the careful distinction, made by both Köhler and Wallach and ourselves, between displacement and fading effects—both theories predict that the latter effect should be maximal at coincidence.

(*b*) It is true, as Smith claims, that the statistical theory could not explain a movement of *T*-contours *toward I*-contours. But Köhler and Wallach do *not* mention this explicitly, as he further claims, citing p. 297; they describe a quite different situation. "If the *I*-object is an oblong, and if two *T*-squares are shown within the affected area, the distance *between these squares* is shortened" (2, pp. 296–7, italics mine). Note that it is the distance between the two *T*-squares, both retreating from opposite *I*-contours, that is decreased. A careful reading of the various examples in the following paragraph indicates that all the effects described are cases of displacement of *T*-contours away from *I*-contours—indeed, had this not been the case, Köhler and Wallach would have been obliged to count them as evidence contrary to their satiation theory. Smith concludes that he "has on file protocols of several rigorous observations by himself and by others in which this anomalous effect showed itself." Such observations would be of great theoretical significance and by all means should be reported in detail.

REFERENCES

1. GIBSON, J. J. Adaptation, after-effect, and contrast in the perception of curved lines. *J. exp. Psychol.*, 1933, **16**, 1–31.
2. KÖHLER, W., & WALLACH, H. Figural after-effects: an investigation of visual processes. *Proc. Amer. phil. Soc.*, 1944, **88**, 269–357.
3. OSGOOD, C. E., & HEYER, A. W., JR. A new interpretation of figural after-effects. *Psychol. Rev.*, 1952, **59**, 98–118.
4. SMITH, K. R. The satiational theory of the figural after-effect. *Amer. J. Psychol.*, 1948, **61**, 282–286.
5. SMITH, K. R. The statistical theory of the figural after-effect. *Psychol. Rev.*, 1952, **59**, 401–402.

[MS. received October 13, 1952]

Psychological Review
Vol. 60, No. 3, 1953

THE LINEAR OPERATOR OF BUSH AND MOSTELLER

RAYMOND H. BURROS
University of Illinois

The purpose of this paper is to correct a semantic misunderstanding about the two parameters in the linear operator of Bush and Mosteller (2). Their basic equation was written as

$$Qp = p + a(1 - p) - bp,$$

where a and b are parameters, p is the probability of response in a specified interval of time, and Qp is the probability of response in the next interval of time. The statement of Bush and Mosteller, recently criticized by this writer (1), read as follows: "To maintain the probability between 0 and 1, the parameters a and b must also lie between 0 and 1" (2, p. 315).

It is likely that some readers interpreted this sentence correctly. Its lack of complete exactness, however, led this writer (1) to misunderstand Bush and Mosteller. The misinterpretation was this: If a and b are chosen outside of the open interval $(0, 1)$, then for *no* value of p in the closed interval is it possible to calculate a value of Qp in the open interval. Taken as it stands, this statement is clearly false, as the writer demonstrated (1). Bush and Mosteller, however, did not mean this.[1] What they did mean is the following: A necessary condition for the proper restriction of Qp $(0 \leq Qp \leq 1)$ for *all* admissible values of p $(0 \leq p \leq 1)$ is that *both* a and b are restricted $(0 \leq a \leq 1,\ 0 \leq b \leq 1)$. The proof follows from the basic equation, here written as

$$Qp = a(1 - p) + (1 - b)p.$$

Since this is assumed for all admissible values of p, set $p=0$, whence $0 \leq a \leq 1$; set $p=1$, whence $0 \leq 1-b \leq 1$, and thus $0 \leq b \leq 1$.

The corresponding sufficiency theorem may now be stated: A sufficient condition for the proper restriction of Qp $(0 \leq Qp \leq 1)$ is the restriction of p, a, and b $(0 \leq p \leq 1, 0 \leq a \leq 1, 0 \leq b \leq 1)$. A derivation based upon a closely related proof by R. D. Luce[2] is as follows: Since $0 \leq a \leq 1$, and $0 \leq 1-p \leq 1$, therefore $0 \leq a(1-p) \leq 1-p$. Since $0 \leq 1-b \leq 1$, and $0 \leq p \leq 1$, therefore $0 \leq (1-b)p \leq p$. Therefore, by addition, $0 \leq Qp \leq 1$.

The reader should note that the necessity theorem is not the converse of the sufficiency theorem, since the restriction of p to the closed interval $(0, 1)$ is postulated, not derived, in both. This lack of a converse relationship between the two theorems was a contributing cause to the writer's misunderstanding.

If $0 \leq a \leq 1$ and $0 \leq b \leq 1$, then $-1 \leq 1-a-b \leq 1$. When Bush and Mosteller applied their basic theory to certain problems in operant conditioning, however, they were implicitly requiring more stringent restrictions on a and b.

In the first place, there is the restriction that $0 \leq a+b \leq 1$. Consider their Equation 4 (2, p. 316) which gives $Q^n p$ as a function of n. It is not necessary for this function to be continuous in n. But later on their use of the calculus does imply this continuity. Therefore, whenever the calculus is used, $1-a-b$ is nonnegative.

[1] Bush and Mosteller. Personal communication. 1952.

[2] R. D. Luce. Personal communication. 1952.

Since this quantity is already restricted to the closed interval $(-1, 1)$, therefore $0 \leq 1-a-b \leq 1$, and thus $0 \leq a+b \leq 1$.

In the second place, the calculus implies that both a and b (and their sum) must be very small compared to 1. Bush and Mosteller (**2**, p. 317) state that

$$\Delta p = a(1 - p) - bp.$$

Since the calculus implies continuity in n, therefore, as Δn approaches zero so does Δp. Since this is true for all permissible values of p, set $p=0$, whence a approaches zero. Set $p=1$, whence $-b$ (and thus b) approaches zero; and consequently so does $a+b$.

The values of a and b obtained to date from curve fitting are quite consistent with these conclusions. Bush and Mosteller computed the following values: $a=0.014$, $b=0.026$ (**2**, p. 320).

The reader should note that so far we have been concerned almost entirely with mathematical matters. What we have deduced is logically true, whether or not the theory is empirically true. Can we, however, rule out in advance the possibility of finding values of the parameters a and b outside the closed interval $(0, 1)$? The writer believes that we should not dismiss the possibility. Bush and Mosteller (**2**, p. 314) derived the linear operator as a "first approximation." It is possible, however, that a more adequate expression for the operator may be nonlinear. If so, a linear approximation may still be quite adequate over a subinterval of the admissible range of p, but not over the entire closed interval $(0, 1)$. In this case, the restrictions on a and b need not hold at all, and yet Qp may still be properly restricted. We have seen that the values of a and b reported to date by Bush and Mosteller are quite properly restricted and so is their sum. If, however, empirical values for a and b are ever obtained outside their proper range, then *one* linear operator will not be an adequate approximation over the entire admissible range of p (for that research at least).

Unless experimental evidence forces us to abandon the linear operator, its simplicity is a strong argument for its continued use. As long as the linear operator is used, especially in conjunction with the calculus, the proper restrictions upon the ranges of the two parameters will be basic to any extension of the theory. That is the justification for the detailed study of this problem.

REFERENCES

1. BURROS, R. H. Some criticisms of "A mathematical model for simple learning." *Psychol. Rev.*, 1952, **59**, 234–236.
2. BUSH, R. R., & MOSTELLER, F. A mathematical model for simple learning. *Psychol. Rev.*, 1951, **58**, 313–323.

[MS. received October 16, 1952]

Psychological Review
Vol. 60, No. 3, 1953

ON A DEFINITION OF CULTURE

MORTIMER BROWN
University of North Carolina

In *Learning Theory and Culture* (1, p. 385) a definition of culture is developed and presented as, "'Culture' = '$\hat{y}$ (x learns y from z and $x \neq z$)' by df. (to be read 'the class of values of the variable y such that x learns y from z, and x is not identical with z')." This definition is arrived at by examining seven fundamental assumptions of Murdock (2) and discarding four of them while retaining two.[1]

From their text we learn that culture thus defined can be read as: culture is defined as the class of responses of any hominid individual learned from any other hominid individual. Culture thus defined appears to be an area of investigation for psychologists (as scientists who are interested in studying the behavior of individuals as such) rather than anthropologists or sociologists whose interest in studying behavior is focused more upon the behavior of groups. Furthermore, the process itself (x learns y from z and $x \neq z$) by which culture ($\hat{y}$) is mediated is the area of psychological investigation which goes by the general name of learning theory and more specifically by the name of human learning.

A stated reason for Moore and Lewis' interest in framing a definition of culture is that it should ". . . have maximum utility in facilitating cross-disciplinary cooperation between psychology, anthropology, and sociology" (1, p. 384). However, by their definition they have cast out any utility in involving anthropologists and sociologists and instead have implicitly indicated a need for cooperation between psychologists and educators whose main job, it may be said, is to facilitate the process whereby (x learns y from z and $x \neq z$).

In order to bring the anthropologists and sociologists back into the picture it seems necessary to involve, in some way, the concept of the "group" as part of the definition. This may be done as follows: "culture = $\hat{y}$ (x learns y from z and $x \neq z$) and where x and z are both members of some homogeneous group." Homogeneous group may be defined as a collection of individuals (x_1, x_2, . . . x_{n-1}, x_n) who already share certain common characteristics. These "common characteristics" may be specifically delimited in anthropological and sociological terms as indicated by findings from research in group membership.

Only by paying some respect to the "group nature" of culture can the anthropologists and sociologists be included in cross-disciplinary cooperation. That they should be included is certainly indicated by dint of their previous work in the field.

[1] Moore and Lewis consider that Murdock's fourth assumption is subsumed under his first, thus making a total of six to deal with rather than seven.

REFERENCES

1. Moore, O. K., & Lewis, D. J. Learning theory and culture. *Psychol. Rev.*, 1952, 59, 380–388.
2. Murdock, G. P. Uniformities in culture. *Amer. sociol. Rev.*, 1940, 5, 361–369.

[MS. received October 6, 1952]

THE CIRCUMNAVIGATION OF COGNITION

BENBOW F. RITCHIE

University of California

Columbus Day is an occasion in this country for celebrating the belief that the earth is round. Since 1492 this idea has caught on so well that it is now a part of the public school curriculum. Opposition to it has virtually disappeared. Today, however, certain new ideas in modern science suggest that we may have been too hasty in our judgment and that this belief may be quite misleading if not actually false. Now by "modern science" I do not mean what you think. I mean, instead, the new methods of "theory construction" as they are called, devised by psychologists.

These methods were devised, of course, to deal with specific psychological problems, but their use need not and indeed should not be limited to these problems. To my knowledge the present paper is the first to apply these methods to problems outside the social sciences. The problem we have chosen is the problem of the earth's shape. Is it round or is it flat? The analysis we have chosen is one recently used by Kendler (4) in his discussion of a similar problem in psychology.

What is the Shape of the Earth?

Geographers have disputed about the shape of the earth since Pythagoras first suggested that it was round rather than flat. This is certainly not the place to review all the arguments, but there is one argument which we must discuss. I refer to the argument based upon what is called "the phenomenon of circumnavigation." By this the ball theorists, as they are called, mean that explorers who set out from some place and keep sailing in a constant direction, eventually return to the place from whence they started. The results of the explorations of Magellan (1), Drake (3) and Captain Cook (8) are all illustrations of this phenomenon. The ball theorists claim that these results contradict the basic assumptions of the disk theory, and they surely seem to, at first sight. But before we decide let us consider the replies which the disk theorists have made to this argument.

Some disk theorists (5) reply by demonstrating that, no matter what the facts appear to be, circumnavigation is impossible. This demonstration is based upon an analysis of the word "to navigate." How do we know, say these theorists, when navigation has occurred? We can only know this if the navigator has moved from one place to another, in short, the empirical meaning of the word means to go to *another* place. Thus the very notion of "circumnavigation" is contradictory since it means to navigate to one's starting place. In this sense the phenomenon is impossible.

Other disk theorists (7) admit that circumnavigation is possible, but seriously doubt that it ever occurs. The fact that a few explorers have "circumnavigated," they say, is given all too much importance. Consider instead the many, many, explorers who have set out to circumnavigate and have failed. Thus the few cases of so-called "circumnavigation," they say, might easily be expected simply on the basis of chance.

There are other disk theorists (6) who admit that circumnavigation does occur, but think that it is a mighty poor way to travel. They point, for example, to the great numbers of travelers who have successfully returned home

by retracing their original route. Thus they demonstrate that the chances of safe return are much greater by this method than by the method of circumnavigation.

The issue, say other disk theorists (2, 9), is not a theoretical one at all. Of course, "circumnavigation" in *some* sense occurs. But in *what* sense is the crucial question. Only when we have discovered all the factors that produce circumnavigation, will we be able to answer this question. And when we have done this, there will be no issue left for theoretical dispute. The facts will have provided the answer.

Finally there are those who might be called "the semantical disk theorists." They say that the controversy results from the use of words. Its solution consists in recognizing that what ball theorists mean by the word "round" is what everyone else means by the word "flat." Once the appropriate word substitutions are made, the problem is resolved. There also is a group of "semantical ball theorists" who apply the same kind of analysis to the problem. They conclude that what the disk theorists mean by "flat" is what everyone else means by "round." The only problem that remains is to decide which "semantical analysis" is the correct one.

So much for the present status of the controversy concerning the shape of the earth. Can the controversy be settled? Certainly there is little hope of either side giving in. What, then, is to be done? Perhaps it is time to apply methodological analysis. Kendler (4) reported great success following his use of such an analysis. He applied it to a problem concerning the nature of learning about which there had been a long and apparently irreconcilable controversy. Following a single application of methodological analysis the controversy was resolved. Because of the remarkable success of this approach to *that* problem we shall employ the same analysis to the problem of the earth's shape.

THE QUESTION AND ITS ANALYSIS

Present-day philosophy of science has devised criteria for discriminating questions that are meaningless from those that are not. So, whenever a question is posed that no one is able to answer, it is time to ask whether the question is answerable. "By application of methodological analysis," says Kendler (4, p. 269), "it is possible to demonstrate that certain problems are not resolvable, not because they are too profound, but rather because the questions they raise cannot be properly answered."[1] If the question can be shown to be a pseudo-question, then all sensible persons will refrain from asking it, and our inquiry can be directed to more fruitful problems. It is the purpose of the present analysis to show that the question, "What is the earth's shape?" is such a pseudo-question, and so should not be asked.

Now of course most geographers not only regard this as a sensible question, but also believe that an answer to it is crucial to an understanding of geography. On the other hand, as we have seen, empirical evidence refuses to provide us with an answer. Consider, for example, the results from various balloon ascensions made by geographers seeking an answer to this question. When they came down and described what they saw from aloft, the descriptions of the ball theorist and the disk theorist were alike in every detail. There is only one difference between them. One describes the earth's surface as round, the other as flat. How is this possible? Methodological analysis states that such a paradox arises whenever the question posed is a

[1] Unless specifically noted, all further quotations will be from Kendler's paper (4).

pseudo-question. This is expressed in one of the fundamental principles of methodological analysis.

> If comparable data are employed to support diverse answers to the same question, then the major source of difficulty lies not in the seemingly opposed answers but, rather, in the question itself.

Now how do these conflicting notions about the shape of the earth arise? When we read the theoretical papers of various geographers, these notions appear to be basic to the theories presented. But are they really? Now, no matter how convinced a geographer may be that his notions about the earth's shape are essential to his *thinking*, these notions may be quite external to his *theory*. At this point it may be helpful to introduce another principle of modern methodology. According to this second principle it is essential to distinguish between a scientist's *thinking* and his *theory*. It was formerly believed that a scientist's thinking produced his theory, and as a result his theory represented his thinking. But this is all wrong. It is based upon a prescientific notion of causation, and so is rejected by modern methodology. In its place we have the sharp distinction between thinking and theory. So, although a geographer may think a great deal about the shape of the earth, his theory need not and perhaps should not make any reference to the earth's shape. At this point the reader may find this distinction between thinking and theory puzzling. However, when we see what is considered "theory" by modern methodology, the distinction should become obvious. And to this matter we now turn.

It has been suggested that notions about the earth's shape may be external to geographical theory. To decide this question, we must review what are called "the structural requirements" of a geographical theory.

The geographer is concerned with stating in as precise a way as he can where things are. This task has two aspects: (*a*) the stating of the location of some given thing or group of things, and (*b*) the description of the thing or group of things in a given location. Now in order to do this the geographer must travel from place to place noting first what is in each place and second how he got to each place. Thus his empirical "first-order laws" as they are called take the following form:

> If I start from place A, and go a certain distance in such and such direction, then I will reach B.

Such a first-order empirical law describes the relation between the independent variables of starting place, direction, and distance, and the dependent variable of terminal place, which results when the antecedent conditions specified by the independent variables, are satisfied. So far, the geographer has no need of theory. If he wishes, he can merely make a list of all the empirical laws discovered in his travels and do without theory. But the geographer, if he travels enough, will discover two remarkable things which may lead him to begin theorizing.

First, he will discover that place B can be reached from a variety of starting places. If he is in a methodological mood, he may express this by saying that the dependent variable is a function of several independent variables. Secondly, he is likely to observe that different terminal places can be reached from the same starting place. These two discoveries lead the geographer to construct or erect a theory. He does this by making a map on which place B, as well as many other places, is represented. The map shows how it is possible to get to B from many of these

places, and also how it is possible to get to many of these places from place B. Such a map, speaking methodologically of course, "bridges the gap existing between the independent and dependent variables."

The geographer with a methodological orientation prefers such maps, which he calls "intervening variables," to a list of directions or rules for getting from one place to another. As he puts it, he would rather create such a theory than "treat separately the relationship each independent variable bears to many dependent variables," and vice versa. It is, of course, important to understand that the intervening variable is not discovered by the geographer. Not at all. It is invented or constructed by him and "this intellectual construction," as he will tell you, "has as its aim the economical description of the known empirical relationships and the prediction of new phenomena." Once you have grasped these essentials of theory construction you are in a position to understand "the structural requirements" of a geographical theory. Such a map, or theoretical erection, must, if it is not to collapse, be anchored to the antecedent independent variables on the one side, and to the consequent dependent variables on the other. Any map which is not so anchored is useless for guiding us anywhere, and so is said to be without operational meaning. Thus, the structural requirements of a theory state in a very methodological way the conditions which ensure that the theory has operational meaning.

It is clear from all this that the structural requirements of a geographical theory include no references to the shape of the earth. In this sense, at least, such statements are external to theory. But modern methodology reveals an even deeper sense in which this is true. Consider for a moment the first-order empirical laws of a ball and a disk theorist. Will there be any differences in these laws? No, for the consequences of going a certain distance in a certain direction from a given starting place will be the same for both. Since a map is merely a shorthand description of such empirical laws there can be no *operational* differences between the maps of two "opposed" theorists. Thus in a deeper methodological sense statements about the shape of the earth are external to geographical theory.

But what, the reader may ask, am I talking about when I say that the earth is not flat? The methodological answer to this question is simple and direct. Nothing! Such statements, the methodologist will tell you, "represent secondary and unnecessary elaborations about the meaning of these intervening variables." You would never make such completely pseudo-statements if you remembered that these intervening variables serve as economical devices to "order" the relations expressed in our first-order laws. These maps, he will go on, "are *shorthand descriptions* and nothing more. . . . The only *meaning* possessed by these intervening variables is their relationship to both the independent and dependent variables. Because this point has been ignored, an immense amount of confusion concerning the 'real meaning' of these intervening variables exists."

But why, one may ask, has such an obvious point been so persistently ignored? The reason, says modern methodology, is the "fallacy of reification or hypostatization." This fallacy consists in regarding certain words as names of things or entities when they aren't. Let's begin by assuming we know what is meant by the words "thing" and "entity." Without such an assumption it is very difficult to make this fallacy understood.

Any *thing* can be given a proper name like "Julius Caesar" or "53A270," and

can also be given a class name like "Roman general" or "Ford sedan." Now, although all class names in English are nouns, not all English nouns are class names. This is most easily illustrated with slang expressions like: "He threw a tantrum," or "She copped a gander." In such cases it is clearly silly to ask where the tantrum is that was thrown, or the gander that was copped. The reason why it is silly is that the nouns "tantrum" and "gander" have no meaning apart from these phrases in which they appear. The fallacy of reification is committed when you regard such a noun as a class name referring to things or entities.

Now as we have seen, a geographer's duties consist in traveling, recording observations on a map. This whole complex process is called "mapping the earth." The geographer commits the fallacy when he thinks of the word "earth" as having some meaning apart from this phrase. He then regards the word "earth" as a class name and imagines that it refers to some thing or entity. It is thus, modern methodology makes clear, that the fallacy of reification creates the problem of the earth's shape. The realization that the word "earth" does not refer to a thing or entity, disposes of the problem.

The Use and Abuse of Intuitive Models

Hume recommended that nonsense, when discovered, be committed to the flames. In his view it could contain "nothing but sophistry and illusion." Modern methodology, however, is not so reckless with the products of human creation. Although, as we have seen, statements about the shape of the earth are nonsense, we should not conclude from this that such statements are worthless. Far from it. They serve to help the geographer in his construction of what is called an "intuitive model." This model serves as a *"thinking aid"* leading to the invention of theoretical constructs and intervening variables. Some geographers, for reasons which are not yet fully understood, get more help from thinking of the earth as flat, others are helped more by thinking of it as round. "It would be hazardous, as well as somewhat presumptuous, for any theorist to insist that every theorist think in his style."

The failure of Hume and others to recognize the usefulness of such nonsense was due to their misunderstanding of the relation between thinking and theory. As we have pointed out, modern methodology makes a sharp distinction between "the personal thought processes leading to the invention of theoretical constructs and the operational meanings" of these constructs.

But the fact that such meaningless statements form the core of scientific thinking should not mislead the reader into thinking that such statements are capable of being either true or false. Modern methodology insists that the decision between various such intuitive models "is in the last analysis a decision having no *truth character*. That is, in spite of the fact that the choice of a model may, and usually does, influence both experimentation and theorizing, the *choice itself* cannot be evaluated as being right or wrong. It is a matter purely of personal taste. The most we can do is to attempt, in a sincere and conscientious manner, to understand the implications of such decisions, but we should not be led astray by believing we can experimentally test their validity. . . ."

Summary

We have almost completed the "circumnavigation of cognition." One further methodological homily will serve to end the trip. Henry Fielding in *Tom Jones* has this to say:

The only supernatural agents which can in any manner be allowed to us moderns, are ghosts; but of these I would advise an author to be extremely sparing. These are indeed, like arsenic, and other dangerous drugs in physic, to be used with the utmost caution; nor would I advise the introduction of them at all in those works, or by those authors, to which, or to whom, a horse-laugh in the reader would be any great prejudice or mortification.

REFERENCES

1. Blodgett, H. C. The effect of introduction of reward upon the maze performance of rats. *Univer. Calif. Publ. Psychol.*, 1928, **4**, 114–134.
2. Brogden, W. J. Some theoretical considerations of learning. *Psychol. Rev.*, 1951, **58**, 224–229.
3. Buxton, C. E. Latent learning and the goal gradient hypothesis. *Contrib. psychol. Theor.*, 1940, **2**, No. 6. 75 pp.
4. Kendler, H. H. "What is learned?"—A theoretical blind alley. *Psychol. Rev.*, 1952, **59**, 269–277.
5. McGeoch, J. A. *The psychology of human learning.* New York: Longmans Green, 1942.
6. Meehl, P. E., & MacCorquodale, K. A further study of latent learning in the T-maze. *J. comp. physiol. Psychol.*, 1948, **41**, 372–396.
7. Miller, N. E. Comments on multiple-process conceptions of learning. *Psychol. Rev.*, 1951, **58**, 375–381.
8. Seward, J. P. An experimental analysis of latent learning. *J. exp. Psychol.*, 1949, **39**, 177–186.
9. Skinner, B. F. Are theories of learning necessary? *Psychol. Rev.*, 1950, **57**, 193–216.

[MS. received for early publication January 22, 1953]

VOL. 60, NO. 4 JULY, 1953

THE PSYCHOLOGICAL REVIEW

THE SIZE–DISTANCE INVARIANCE HYPOTHESIS

F. P. KILPATRICK AND W. H. ITTELSON

Princeton University

Extremely common in the literature of visual perception are explanations or descriptions of the relationship between size, distance, and visual angle which resort to the use of mathematical expressions which are special cases of the proportions existing in similar triangles. Such applications of Euclidean plane geometry to psychological relationships, with their consequent, often uncritical, mixture of physical and psychological variables, offer attractively simple solutions. However, the value of these solutions rests at bottom on the degree to which they actually are consistent with the observed relationships. Consequently, a careful scrutiny in relation to experimental evidence of the hypothesis from which such mathematical expressions are derived would seem to be in order.

The basic expression follows directly from an elementary theorem of plane geometry. The proportions between corresponding parts of similar triangles shown in Fig. 1 can be expressed as

$$\frac{a_1}{b_1} = \frac{a_2}{b_2}.$$

FIG. 1.

It follows directly from this that specifying α (Fig. 2) determines the relationship between a, b, and β. Specifically this relationship is given by

$$\tan \alpha = \frac{a \sin \beta}{b - a \cos \beta}. \qquad [1]$$

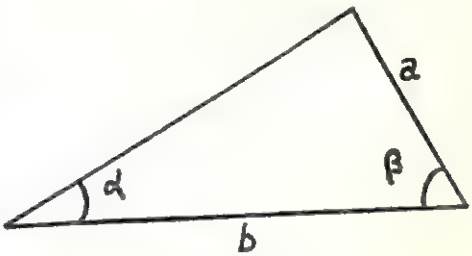


FIG. 2.

We can apply this simple geometry to physical objects in physical space, if we confine ourselves to sizes and distances which are neither too small nor too large.

Let S be the physical size of an object (Fig. 3) at a physical distance D from some reference point O, let α be the angle subtended by S at the point O, and let δ be the deviation of S from the vertical, or the *inclination* of S.

Equation [1] becomes

$$\tan \alpha = \frac{S \cos \delta}{D - S \sin \delta}. \qquad [2]$$

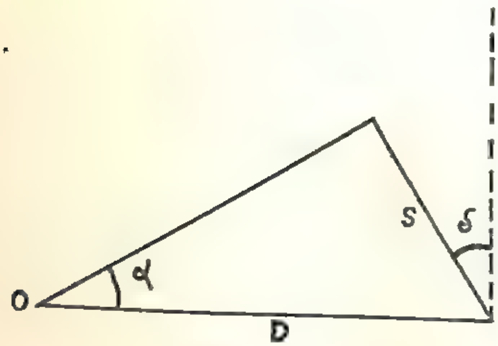


FIG. 3.

If we limit ourselves to angles sufficiently small so that $\tan \alpha = \alpha$, and consider as a special case the condition $\delta = O$ (Fig. 4), equation [2] reduces to

$$\alpha = \frac{S}{D}. \qquad [3]$$

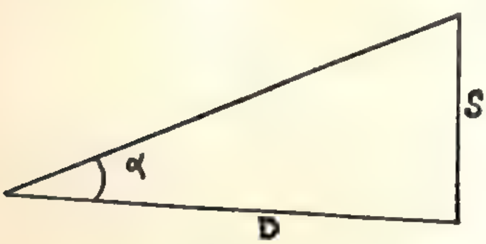


FIG. 4.

In this paper we shall consider the implications only of equation [3], bearing in mind at all times that it is only a special case of equation [2], which is in turn a special case of the three-dimensional relationship, including the most general case in which S is a curved surface.

From equation [3] certain simple conclusions may be drawn.

a. When the angle is given, size and distance vary proportionally.

b. When angle and distance are given, size is uniquely defined.

c. When angle and size are given, distance is uniquely defined.

d. When size is given, angle and distance vary inversely.

e. When distance is given, angle and size vary proportionally.

This simple geometry of the right triangle is of interest to the student of space perception because the apparent sizes and apparent distances of objects seem to conform to similar or identical rules. Indeed, in many cases equation [3] seems quite accurately to describe the relationship between apparent size, apparent distance, and visual angle. This fact has led some psychologists to apply this equation to the description of psychological events. In many instances, the resulting equations injudiciously mingle physical and psychological quantities. An example is

$$\text{visual angle} = \frac{\text{apparent size}}{\text{physical distance}}.$$

Another familiar example which leans more heavily on psychological quantities is

$$\text{retinal size} = K \frac{\text{apparent size}}{\text{apparent distance}}.$$

However stated, one assumption is central to all such applications of simple geometry to perceived space: *a retinal projection or visual angle of given size determines a unique ratio of apparent size to apparent distance.* This is the Size-Distance Invariance Hypothesis.

This hypothesis has been utilized by psychologists in two ways. First, as *explanation*, some form of invariance hypothesis is central to many theories of the perceptual constancies. As thus used, the most general statement of the invariance hypothesis holds that constancy depends on the existence of two processes which vary reciprocally so that the resultant product always re-

mains the same. This type of explanation as related to the size-distance problem was first proposed by Wheatstone; it is central in the Gestalt theory of constancy, implied in much of Boring's writing, and again crucial and explicit in Gibson's theory, to mention a few important examples (cf. the discussion in [11]). Also, it has found expression in recent issues of this journal in the writings of Schlosberg (16), who employed it to explain experiments in the visual perception of size and distance, including a number of the Ames demonstrations, and of Gelinsky (8), who used it as a basis for a derivation and quantitative formulation of "a general unifying law of visual space perception . . ." (p. 475).

In this paper we shall not be directly concerned with the use of the invariance hypothesis as an explanatory concept in perceptual theorizing, but rather with its second major use, that of *description*. We choose to concentrate on this latter aspect because it is more fundamental; the adequacy of a formulation as explanation depends directly on its adequacy as description. Is the invariance hypothesis adequate to describe the observed relationships between apparent size, apparent distance, and visual angle? The existing evidence will be examined and a new experiment reported in an effort to answer this question.

That the size-distance invariance hypothesis does describe this relationship in many instances is well attested. No attempt will be made here systematically to survey the literature on this subject but rather simply to indicate the general nature of the supporting evidence. Three cases may profitably be distinguished.

a. Visual angle constant. The invariance hypothesis demands that apparent size and apparent distance vary proportionally, and indeed this is the most commonly reported result. Duke-Elder has summarized current evidence of this case by noting that "If we estimate the distance to be much greater than it is, we consider that if the object at this distance is so big, it must be very big indeed. Conversely, if an object seems nearer than it is, it seems to be smaller than it actually is" (6, p. 1071). This case is also the one with which Emmert's law is concerned, whose usual interpretation is in support of the invariance hypothesis.

b. Apparent distance constant. The invariance hypothesis demands that apparent size vary directly with visual angle, and again this is the most commonly reported result. This case is, in fact, closely related to the finding of most size-constancy work, that the proper estimation of size is dependent upon the proper estimation of distance.

c. Apparent size constant. The invariance hypothesis demands that apparent distance vary inversely with visual angle. This general finding has been repeatedly verified in a variety of cases.

It is clear, then, that considerable experimental evidence can be mustered in support of the invariance hypothesis, and indeed most writers on the problem of the relation between apparent size and apparent distance would seem to be in general agreement with this hypothesis. The prevailing view can be summarized in Sanford's words written half a century ago, ". . . size and distance are mutually determining. If the apparent distance is constant, the apparent size of the objects changes directly with the size of the retinal image; while if the apparent size is constant, the apparent distance changes inversely with the image. These are facts of very common observation" (15, p. 198).

Possibly the case could be allowed to rest there, were it not for the fact that there is also considerable experimental

evidence in support of the proposition that no aspect of perception can be a simple and invariant function of a few variables. This proposition is derived from the more general theory that the perceptual process is one which always and necessarily includes all aspects of the functioning of the organism (1, 4, 5, 11, 12, 13). To find that the size-distance relationship is an invariant one as so commonly reported would constitute an important exception to that theory. Consequently, it would appear to be of considerable systematic importance to note that the findings with respect to the invariance hypothesis are not entirely one-sided. Although the contradictory evidence is scattered and possibly not conclusive, it stems from a variety of experimental approaches.

a. Varying accommodation while looking at a fixed object. The relation between accommodation and apparent size under these conditions is well established. Anyone can verify for himself the truth of Hering's report that "when changing from a monocular fixation of a distant object to viewing one close by, the distant object appears to diminish in size. Conversely the size of the nearer object increases when accommodating for the farther" (10, p. 172). The relation of this effect to apparent distance is by no means so clearly established. This problem has never been systematically investigated, but it has been mentioned by many writers, who, although they are not in complete agreement, would seem to support von Kries in stating:

When we gaze at any object with one eye screened, it is easy to notice that its apparent size varies with the state of accommodation. Every exertion of accommodation is accompanied by an apparent reduction in size, and every relaxation by an apparent magnification. The simplest explanation of this well-known and easily observed phenomenon would be to suppose that as the result of accommodation for near vision the object appears to be at the distance that ordinarily corresponds to this accommodation, that is, too near; and hence, as long as it subtended the same visual angle, its absolute size would appear to be less than it was. The trouble about this explanation is that, as Donders long ago rightly pointed out as something remarkable, this is not what actually happens. The object does appear to become smaller when accommodation is exerted, but it by no means appears to come nearer at the same time; on the contrary, it appears to recede. Thus, with a constant visual angle, we see here apparent size and apparent distance vary in the opposite sense (17, p. 389).

This is clearly and explicitly in contradiction to the invariance hypothesis.

It should, however, be pointed out that this evidence, as well as that in the following section, is complicated by the possibility that the physical size of the retinal projection varies with changes of accommodation. (See, for example, discussions in Neumueller [14] and Duke-Elder [6]).

b. Accommodative micropsia due to partial paralysis of accommodation. Partial paralysis of accommodation, most commonly by means of atropine, is accompanied by a reduction in apparent size, or micropsia. While this fact is well established, the relation to apparent distance again is not completely clear. Duke-Elder implicitly attributes this effect to the invariance hypothesis when he writes that ". . . in partial paralysis of accommodation . . . we volitionally expend a greater effort to see an object distinctly, . . . and we thus imagine the object to be quite near; but since the retinal image is of the same size, we think the object must be smaller than it is" (6, p. 1072). Hering, however, reported findings of Aubert which specifically refute this explanation. "Aubert after instillation

of 1/500 grain atropine, saw letters with the atropinized eye at a distance of 20 feet, apparently half the size as seen with the other. . . . In spite of this, strange to say, the apparently smaller objects did not seem nearer but farther than in the sound eye" (10, p. 172). This report is clearly in contradiction to the invariance hypothesis.

c. Visual tau effect. "If three discrete points of light are successively presented, and the distance between the points of light is equal, but the time interval between the first and second points of light is greater (lesser) than the time interval between the second and third, then the perceived space between the first and second will be greater (lesser) than the perceived space between the second and third points of light" (7, p. 483). Contrary to what one might expect in terms of the invariance hypothesis, there is no evidence that this well-attested phenomenon is accompanied by any apparent distance change. Of course, it might be argued that the perceived space between the successively presented lights is not "size" in the usual sense, as the lights which set the "space limits" do not stimulate the retina simultaneously. Nevertheless, pending further experimentation, the visual tau effect might reasonably be considered an exception to the invariance hypothesis.

d. Subjective factors. Some recent evidence seems to indicate that the apparent size of an object may be influenced by subjective factors such as the importance of the object to the observer (cf. discussions in [2, 3]). While such evidence is as yet inconclusive, if it should be substantiated, it would constitute an important exception to the invariance hypothesis.

e. Statistical considerations. Much if not all of the evidence which tends to support the invariance hypothesis is statistical in nature. That is, it states that the invariance hypothesis describes the mean of a large number of observations in a large number of cases. But as usually expressed, the invariance hypothesis is not a statistical statement; it is an equation which, within the limits of accuracy of measurement, must apply to the individual case. Therefore, even if the invariance hypothesis always adequately described the mean performance (which, as we have seen, it probably does not), it would still have to be modified drastically before it would be acceptable as a description of psychological events.

It is clear, then, that there exists a body of evidence which, although it approaches the problem obliquely, is yet relevant and contradictory to the invariance hypothesis. The evidence at the very least suggests that this hypothesis may not be adequate to describe the relationship between apparent size, apparent distance, and visual angle in all cases. There is need, nevertheless, for an experiment specifically designed to test the applicability and generality of the invariance hypothesis. In the experiment reported below visual angle was kept constant, apparent distance was influenced by over-lay indications, and apparent size by the size of the background. Reports were obtained of changes in apparent size and apparent distance and were analyzed as supporting or contradicting the invariance hypothesis.

Description of the Experiment [1]

The apparatus employed has been described and illustrated by the authors in a previous publication (13). Briefly, it consisted of a trapezoidal shape cut out of ⅛-in. aluminum, 39 in. long, 47¼ in. and 25 in. high at the long and the short ends, respectively.

[1] The authors are indebted to Charles T. Albert for collecting the data in this experiment.

Three rows of five openings somewhat like the panes in a window were cut out of the trapezoidal shape, and on it were painted shadows like those commonly seen on an ordinary window.

This trapezoid was suspended from the ceiling by two wires so that the sides were vertical and the center of the frame 46 in. above the floor. A black thread passed through the middle opening of the shorter side and was attached to four pulleys so that the thread was horizontal to the floor, 48 in. above the floor, and at right angles to the line of sight. A motor attached to one of the pulleys drove this thread at the rate of about half an inch per second. This apparatus was viewed from a distance of 10 ft. utilizing a headrest which permitted monocular observation only, and which was adjustable so that eye level could be maintained at 46 in. above the floor.

Black drapes were hung behind the trapezoid and between the trapezoid and the observer. In this latter drape, 4 ft. from the viewing point, a rectangular opening was cut through which the trapezoid could be observed. Behind this drape two 15-w. lights in photographic reflectors were mounted so that they gave balanced illumination to both sides of the trapezoid. With the exception of these two lights the room was completely dark.

Two objects were provided which could be attached to the black thread and thereby moved through the trapezoid. These objects were an ordinary playing card (three of diamonds, 2¼ in. × 3½ in.) and a piece of cotton approximately 2 in. in diameter.

Observers were twenty-four undergraduates at Princeton University, none of whom had any previous knowledge of the nature or purpose of this experiment. All of them had, or were corrected to, 20/20 visual acuity.

The trapezoid described above appears to *O* to be a rectangle with the actually far edge of the trapezoid appearing as the near edge of the rectangle and vice versa. Under these conditions, objects carried through the trapezoid in a straight path by the moving thread appear typically to move through an "S"-shaped path in the horizontal plane (see 13). The problem, as related to the study of the invariance hypothesis, is to determine whether this change in apparent distance as the object moves is accompanied by the corresponding change in apparent size which would be predicted from the invariance hypothesis. In order to secure this information, each *O* was presented with three different situations.

1. *Trapezoid alone.* The *O* viewed the trapezoid and was asked to describe its orientation in space and its approximate dimensions.

2. *Trapezoid with moving objects.* Each of the two objects described above was viewed moving through the trapezoid from right to left and also from left to right. The order of presentation was randomized. The *O* viewed each condition twice and was asked to describe any changes in apparent size or distance of the moving object.

3. *Trapezoid with stationary objects.* Two playing cards were suspended from the stationary wire with their inner edges lined up from the viewing point with the outside edges of the trapezoid.

FINDINGS

1. *Trapezoid alone.* All *Os* described the trapezoid as a rectangle tipped sharply sideways with its left edge near and its right edge far. Estimates of its dimensions ranged between 1½ ft. × 2 ft. and 5 ft. × 3 ft., with 18 out of the 24 *Ss* estimating the horizontal dimension to be at least 3 ft.

2. *Trapezoid with moving objects.* In 75 of the 96 trials, definite movement in depth of the test object was

reported, 3 ft. or more in 39 of the trials and less than 2 ft. in only 12 of the trials. Perceived movement in depth was about equally common for the two test objects; in the case of the playing card, such movement was seen in 36 out of 48 trials, while with the cotton it was seen in 39 out of 48 trials. No significant effect of the right-left or left-right orders appeared for either test object, although depth movement was reported a little more often for both objects in the right-left trials. These depth changes, when seen, were quite consistent in length throughout the trials of an individual, and, in general, corresponded to *O*'s earlier estimate of the horizontal dimension of the frame. There can be little doubt, then, that most *O*s in the majority of their trials did see the test object change distance, and that these changes were of sufficient magnitude relative to the viewing distance of 10 ft. that a corresponding change in size should have been perceived in all such cases if the invariance hypothesis were to be substantiated.

a. Playing card. In only 14 of the 37 trials where either size or distance change was seen could the *O*s' reports be interpreted to be in line with the invariance hypothesis. These 14 reports were that when the card approached, it became smaller, and when it moved away, it became larger. In the remaining 23 trials, three quite different size-distance perceptions were reported, all of them distinguished by their lack of accord with the invariance hypothesis. The most common of these deviant observations, reported in 15 trials, was that a marked change in distance was seen without any apparent alteration in size. Another report, occurring in 7 instances, was exactly the reverse of the invariance-hypothesis formulation; when the card approached it became larger, when it receded it became smaller. In the remaining instance no change in distance was reported but the *O* said he saw the card change size.

TABLE 1

REPORTED SIZE-DISTANCE PERCEPTIONS OF THE MOVING OBJECTS

(24 observers, 4 trials each)

Response Category	Number of Trials		
	Card	Cotton	Total
No change in size or distance	11	9	20
Changed in accord with invariance hypothesis	14	15	29
Changed, but *not* in accord with invariance hypotheses*	23	24	47
Total trials	48	48	96

* This category includes three types of size-distance relationships: (*a*) change of size and not distance, (*b*) change of distance and not size, and (*c*) decrease of size with increase in distance or vice versa.

b. Cotton. The results with the moving piece of cotton were substantially the same as those reported for the playing card. The numerical distribution of the findings is reported in the second column of Table 1.

3. *Trapezoid with stationary objects.* Here, also, with the visual angles subtended by the two playing cards the

TABLE 2

REPORTED SIZE-DISTANCE PERCEPTIONS IN THE STATIC SITUATION, RIGHT-HAND CARD STATED IN RELATION TO LEFT

(24 observers, 1 trial each)

Response Category	Number of Trials
Same size and distance	5
In accord with invariance hypothesis:	
Larger and farther	10
Not in accord with invariance hypothesis:	
Larger and nearer	5
Larger and same distance	2
Vacillating between "larger and farther" and "larger and same distance"	2
Total trials	24

same, there is no invariant relationship between reported apparent size and distance. The modal category of response (10 *Os*) is in line with the invariance hypothesis, but almost an equal number of *Os* (9 cases) reported perceptions which could not be so interpreted. The results of this part of the experiment are summarized in Table 2.

DISCUSSION

The size-distance invariance hypothesis states that for a given visual angle there is a unique and constant ratio of apparent size to apparent distance. The results reported above, together with the earlier evidence, clearly indicate that this hypothesis is not adequate to describe this relationship in all cases. The invariance hypothesis therefore loses its status as a powerful explanatory concept and becomes rather a description of results obtained under conditions which have yet completely to be specified.

This conclusion certainly should not be surprising. The invariance hypothesis is commonly expressed in the form

$$\text{visual angle} = \frac{\text{apparent size}}{\text{apparent distance}} \qquad [4]$$

or some equation reducible to this. Such equations represent an attempt to take a relation derived under the axioms of Euclidean plane geometry and to transfer it bodily to the description of psychological events. There is no a priori reason why this should be possible.

When applied to physical objects in physical space, and within certain limiting conditions, the equation

$$\alpha = \frac{\text{physical size}}{\text{physical distance}} \qquad [5]$$

always holds. This equation is useful to the psychologist because it defines the physical limits of the experience of the organism. Whenever the retina has been stimulated along a length describable by the visual angle α, this has always been produced by physical objects at physical distances uniquely specified by the above equation (or more generally by equation [2]). There have never been any exceptions to this rule; there never will be. When equation [4] adequately describes the reported perceptions, this simply means that the perceptions most nearly approximate the physical situations which have been related to all the size-distance experience of the organism, and which can be summarized by equation [5].

When behavior of this type is encountered, it can be labeled *invariance behavior*. The conditions under which invariance behavior can be observed have yet to be determined, but should provide a more satisfactory base line for evaluating experimental results than the usual criterion of constancy. It is, so to speak, what one would expect the organism to do if there were no reason for doing otherwise.

REFERENCES

1. AMES, A., JR. Visual perception and the rotating trapezoidal window. *Psychol. Monogr.*, 1951, **65**, No. 7 (Whole No. 324).
2. BLAKE, R. R., & RAMSEY, G. V. (Eds.) *Perception, an approach to personality.* New York: Ronald, 1951.
3. BRUNER, J. S., & KRECH, D. (Eds.). *Perception and personality.* Durham: Duke Univer. Press, 1949.
4. CANTRIL, H. *The "why" of man's experience.* New York: Macmillan, 1950.
5. CANTRIL, H., AMES, A. JR., HASTORF, A. H., & ITTELSON, W. H. Psychology and scientific research. *Science*, 1949, **110**, 461–464, 491–497, 517–522.
6. DUKE-ELDER, W. S. *Textbook of ophthalmology.* St. Louis: Mosby, 1936.
7. GELDREICH, E. W. A lecture-room demonstration of the visual tau effect. *Amer. J. Psychol.*, 1934, **46**, 483–485.
8. GELINSKY, A. A. Perceived size and distance in visual space. *Psychol. Rev.*, 1951, **58**, 460–482.

9. HELSON, H., & KING, S. M. The tau effect: an example of psychological relativity. *J. exp. Psychol.*, 1931, 14, 202–217.

10. HERING, E. *Spatial sense and movements of the eye.* (Trans. by C. A. Radde.) Baltimore: American Academy of Optometry, 1942.

11. ITTELSON, W. H. The constancies in perceptual theory. *Psychol. Rev.*, 1951, 58, 285–294.

12. KILPATRICK, F. P. (Ed.) *Human behavior from the transactional point of view.* Hanover, N. H.: Institute for Associated Research, 1952.

13. KILPATRICK, F. P., & ITTELSON, W. H. Three demonstrations involving the visual perception of movement. *J. exp. Psychol.*, 1951, 42, 394–402.

14. NEUMUELLER, J. The correction lens. *Amer. J. Optom.*, 1948, 25, 247–261.

15. SANFORD, E. E. *A course in experimental psychology.* Boston: Heath, 1901.

16. SCHLOSBERG, H. A note on depth perception, size constancy, and related topics. *Psychol. Rev.*, 1950, 57, 314–317.

17. VON KRIES, J. Notes. In H. von Helmholtz, *Physiological optics.* (Trans. by J. P. C. Southall.) Optical Society of America, 1925.

[MS. received April 10, 1952]

Psychological Review
Vol. 60, No. 4, 1953

FORMALIZATION OF LANGUAGE SYSTEMS FOR BEHAVIOR THEORY

F. H. GEORGE

University of Bristol

This paper is intended to further the analysis that has already been partially developed (6). It is now hoped to suggest steps of a more positive nature designed to develop adequate language systems for the discussion of behavior theories. It seems to the writer that two of the most pressing needs in modern behavior theory are: (*a*) The development of a more precise language for the discussion of the theory, since much of existing behavior theory is vitiated by an apparent inability to formulate sensible questions. The questions asked prior to experiment, for example, or those which form the basis of theory construction, are, all too often, either meaningless, circular, or insufficiently precise, and this is partly, at least, a function of the language used. (*b*) The development of systematic theories of behavior in molecular terms, as well as in molar terms, appears necessary to remove the vagueness that surrounds molar constructs. This vagueness can be greatly reduced by breaking the constructs down to a molecular level. This paper is wholly concerned with finding a solution to (*a*).

The Development of Formalization

"Formalization brings out clearly to what a large extent problems in science are not empirical questions but questions of linguistic convention. This also helps to eliminate useless controversy." This quotation from a memorandum by Woodger[1] forms the basis for this discussion. Formalization is, broadly, the process of reducing our linguistic systems to more precise logical forms; axioms, definitions, and so on. In this paper it is intended to indicate the desirable development which formalization of behavior theory should take. This approach, it is hoped, will especially be brought to the notice of those narrow experimentalists who repeatedly call for experiment, and decry theory, and continue to stumble through the maze of science erroneously believing themselves to be dealing wholly with facts, and never with linguistics.

It is apparent that there are many symbolic logics of various forms, some of which have already been used in the psychological field, so far with somewhat limited success. Such systems may be premature in a subject which is not clearly developed as a science. Their compact shorthand, although in some form certainly desirable as an ultimate aim, needs to be translated back into longhand for many purposes. Thus our aim is a rigorization of language, or the setting up of a compromise between rigid postulational systems of a narrow kind, such as the propositional calculus, and the vague, inadequate language of ordinary discussion. The principal trouble in existing psychological languages is either (*a*) they sprawl, introduce enormous numbers of new terms, and are vague and lend themselves to meaningless formulations, or (*b*) they are impressively precise and impose a rigidity on our knowledge of behavior that leads to an arbitrariness which allows of no real prediction. More precisely, (*a*) they allow no prediction which is precise, or (*b*) allow precise prediction which is almost wholly false to facts.

[1] Unpublished manuscript.

Ideally it seems that we should have a hierarchy of languages which moves from the very general to the extremely particular and allows us to suit the precision of our language to the precision desired in the discussion. We talk vaguely in ordinary conversation, but become more precise as the dimensions of our topic become more precise, and ultimately we are forced back, as is necessary, to definitions and even to the assumptions underlying our theoretical system. It is certainly true that this last and highest approximation, to which we refer in cases of obscurity, could well be in mathematical terms, or some such rigorous form. This paper, then, will attempt to suggest the way we may fill the gap between our ordinary language and our logical calculi.

In an earlier paper (6) the vagueness accruing at certain points even in rigorous theory, i.e., in the use of logical constructs, undefined terms, philosophical directives, many-meaninged terms, etc. was noticed. This should be borne in mind throughout this paper and will be referred to from time to time.

THEORY CONSTRUCTION

In science, the search for adequate methods of theory construction continues, and psychological theory is especially interested in deriving some relatively precise methods. In this, and subsequent paragraphs, it is intended to outline some of the problems and one way in which they might be solved.

In the first instance we may choose to build our models in at least one of two ways. We start from our experience in either case, but we may find it convenient to assume the existence of a real external world behind experience, which needs to be observed and mapped on various levels of language. This is what Pragmatic Realism aims to do (5, 15). We may, on the other hand, start from experience, without making any assumptions about a real external world; thus our theoretical model comes to consist of observational propositions, and our vague and difficult formulations will be defined operationally. This last method is generally referred to as some brand of Phenomenalism, Positivism, or Operationism (1, 2, 3).

I would submit that there is no obvious reason for not accepting a mixed Operational-Realist system wherein we may regard the Realistic model as being generally adequate, and the use of an operational definition as a form of limiting test where disputes, difficulties, etc. occur—generally at the borders of our knowledge at any given time. The ability to use such operational definitions is a guide to the degree of validity of the theory. The question of setting up criteria for deciding what constitutes a valid operation is a matter for convention, as is the problem of "observation-in-principle." To recapitulate, we are asserting that concepts, or formulations, need not, in general, be defined in a strictly operational form, but that we may conveniently use a realistic model subject to an operational test, wherever possible, and then define our terms in a manner we shall call a contextual definition.

There are now further problems in theory construction to be settled.

"INTENSIONAL" AND "EXTENSIONAL" MODELS

The question now to be considered is the actual method of construction that is best suited to behavior theory. Let us first consider the classical postulational method which has proved to be so successful in mathematics and mathematical science. The theories of Groups, Matrices, Euclidean Geometry, and Boolean Algebra are well-known examples of postulational systems. (It

will be noted that these markedly successful cases are essentially analytic, or conceptual systems.) In psychology Hull has, of course, used a kind of postulational method in the statement of his theory of behavior (9).

The postulational technique itself is not the question which I primarily wish to discuss, as I intend to assume that in some form this is essential. The question is rather: Is the intensionally *defined* system adequate? To put it in other words, we may start from a set of undefined terms and from their derived postulates, and then introduce the various conceptual terms such as learning, perception, motivation, habit, habituation, adaptation, reinforcement, and so on. These terms are not necessarily all to be included in the theory, but some such terms, many-meaninged in character, must certainly arise. The many-meaninged, or multiordinal (having a different meaning on each level of language) types of term must, and will, arise in scientific theory; and they present the central problem. Now again to put the question: are our definitions to be intensional?—i.e., is a class definition adequate? May we, for example, talk of the class of reinforcers, as does Meehl (14)? The answer appears to be that the notion of class is often objectionable and the source of both vagueness and crude approximation. The alternative is an extensional definition, which involves the enumeration of properties, and no question of class definition, although the word "properties" also has its drawbacks, as has been pointed out by Woodger (16).

Let us now compare what I have called intensional and extensional definitions. If class definition is accepted, then we set up a class of some kind, subject to some class criteria; we categorize in fact, and in doing this, differences are overlooked at the expense of similarities. The method is compact and powerful. Consider, for example, the enormous power inherent in the "Boolean Algebra of Classes." The power of the method depends on our ability to categorize without too much distortion. Can this always be done? The answer must be: not always in subjects served by ordinary language without vagueness and arbitrariness. On the other hand, extensional definitions do not allow conciseness, convenient generalization, or a calculus; they do, however, gain in accuracy of statement. The trouble essentially surrounds the question of classification. As soon as we classify, we overlook differences.

The problem may be solved by a compromise. It is literally true that behavior theory needs the benefits of both systems. Thus extensional definitions may be thought of as limiting cases of intensional definitions, when the number of classes tends to infinity. The form of definition to be employed must therefore allow a degree of approximation, suiting the dimensions of the subject discussed. We shall propose, then, that our definitions should start from a glossary definition—a short statement of the common properties involved in the various uses of the term. Then on each occasion that the term is used, it will be contextually defined. The contextual definition may be explicit, and certainly it should be capable of explicit definition. If we now place this common glossary definition, and its set of contextual definitions, in a model of an infinity of levels of language, it will be seen that we have made explicit a simple, and powerful, device for building language systems. The glossary definition, at any level, may be taken as a class definition, if, for example, for quantification purposes, it was necessary to have a precise, if arbitrary, statement. This may be profitable on some dimensions

at some times. At the same time there will exist a general extensional background of levels of propositions with any degree of accuracy of definition possible. A simple example of this usage must suffice. At the level of molar behavior theory a central term is "learning." A considerable amount of discussion takes place without an agreed criterion for learning. The result is confusion, in many cases, not only because two people use different criteria, but because they both, themselves, use the term in various ways. In order to discuss "learning," a start could be made with a glossary definition such as the one from the monograph of Hilgard and Marquis (8), and each time the term was used it would be in this general sense, unless some explicit contextual definition was appended.

Science has, of course, progressed by bringing more and more under less and less, in the sense that certain postulates can be used to subserve more and more diverse and complex propositions. It is this generalization, simplification, and unification that makes science so powerful. However, the price paid by this advance can be too great if it leads to arbitrariness, and this appears to be the present position in behavior theory. Certainly it appears as yet to be impossible to encompass the great variety of behavior in a rigid postulational system. This, at any rate, is an assumption we shall make. Thus explicitly an attempt must be made to build an extensional system which might, as it were, be fixed and made exact, at any level, at any time. In practice this will only be useful if we can suit the fixing to the appropriate dimensions.

Now our criteria for theory construction may be summarized. We must state our undefined, or ostensively defined terms, and in giving our ostensive discussion we must, as far as possible, explicitly state our philosophical directives (i.e., those personal prejudices, etc. that we possess). Next, we must build a set of postulates which will contain words such as "learning," "perception," and so on. Avoidance of these particular terms only throws the burden on some other many-meaninged terms. What has been said does not mean that we are rejecting the rigid postulational methods as used by Hull, and his associates (9, 10); or that we are returning to straightforward general discursive methods such as have been generally used in the work of Hebb, Thorndike, Tolman, etc. The general discursive method is too vague and puts us in a position which leads to discussion about terms, and the linguistic conventions, which are not clear-cut, and this state of affairs seems quite inadequate. The alternative, however, is not at this stage to go to the strictest type of Hull system. This—and this is the principal point made in this article—leads to an arbitrariness which is too narrow for the present development of behavior theory.

The steps that could be taken in learning theory would be to reduce the number of undefined concepts by defining in terms of a small nucleus of undefined terms (necessary in any verbal system), and by setting up a glossary definition for vague, previously undefined concepts, many-meaninged terms, etc. where the "glossary definition" should refer to the common properties of all usages, and then a particular contextual definition given to each usage. If a term is used in, say, two ways that have no common property, then a new term should be introduced. If this pattern is followed explicitly, rather than implicitly, much clarity would result.

It is necessary to bear in mind the distinction made between propositions,

which are part of our strict behavior model, and propositions of the metalanguage, which refer to the propositions of the model. This distinction is implicit in our linguistic theory which is assumed to be on any of a possible infinity of levels. This sort of distinction (13) is necessary to cater for antimonies, and is analagous to the Russell "Theory of Types." With respect to the terminology which we may employ, it is, I believe, true that the Carnap axiom (4) holds and that each subject will develop its own terminology, and will gradually discard the terms that perform no useful function in the subject.

Thus it is proposed that the construction of behavior theory involves first the rigorization of our language systems. We then proceed extensionally, by careful contextual definition of terms, to theorems, which are derived from postulates which refer both to our explicitly stated undefined terms and our glossary definitions. It is worth mentioning that some such language system has recently been developed by Woodger (16) and previously a somewhat similar system was developed by Lesniewski. Woodger aims to avoid the enumeration of properties and the difficulties of class, and he has built an extensional type of calculus which allows discussion to take place and allows degree of accuracy to be involved by devices known as "time-slicing." This is similar to the methods of "indexing" suggested by Korzybski (13), and involves, at the crudest level, the distinction between Walter Scott 1843 and Walter Scott 1844. This can clearly allow of degrees of accuracy, analagous to the infinitesimal calculus, as 1843 can be subdivided into March 1843, March 3rd 1843, and so on indefinitely as we take smaller and smaller "time-slices." This is precisely the sort of graded accuracy on various levels of discourse, or levels of abstraction, that is needed in behavior theory, although we have used the phrase "enumeration of properties," to avoid the notion of definite class. It may now be seen that the precision of glossary and contextual definitions is essentially related to dimensions, i.e., what is adequate as a context definition on one level, may only serve as a glossary definition on another level.

Applications in Behavior Theory

Of all branches of behavior theory, theories of learning and perception are the most philosophically and semantically sophisticated. The fields of abnormal behavior, Personality, and so on, are in a much more primitive state of development and are far from the Carnapian state of throwing out the functionally useless terms. Thus the sort of linguistic systems suggested apply even more obviously to the latter fields, but I believe they are also enormously important to the former. The whole subject of redefinition of psychological terms could well fill many volumes.

The first question is how far may terminology be stretched. The best example extant is perhaps the terminology of the conditional reflex. Recently the Pavlovian terminology of the conditional stimulus, conditional response, unconditional stimulus, unconditional response, generalization, etc. has been extended from the original "meaning" (8, 12) to include more complicated behavior activities. This involves the extension of definitions of some of the terms, and has the advantage of bringing more and more forms of behavior under the same postulational system. However, this is only worth while if the amount of stretching necessary in the definitions is not too considerable; otherwise, it would seem more useful to introduce a new term, and utilize it un-

til such time as the integration may be effected without undue stretch. The actual test must be that of overlap referred to previously. In any event here we may clearly see the use of the contextual definition. If we are discussing generalization, it is important, granted that it has to some degree been stretched, to say precisely, in each instance, what we mean by generalization, and if necessary proceed back to the observations to be carried out to elicit the phenomenon we are naming.

The broader problem of whether conditioning terminology should be used, or not, in the wider context of Instrumental conditioning, in a laudable attempt to bring more of behavior under one concise terminology, can only be decided by attempting to verify the theorems that emanate from the assumptions, i.e., is it possible to carry out the plan without serious distortions and omissions?

Gibson (7) has carried out something of the desired type of analysis for the term "set" with the idea of finding some common meaning for the various usages. He found none which might serve for a glossary definition, but instead, unearthed ambiguities and contradictions. It is precisely this messy uncontrolled growth of terminology that ultimately leads to a lack of clarity and vagueness in the subject that makes prediction impossible. It is clearly essential that further surveys of the kind made by Gibson should be carried out in order to clear up the confusions of the past, and to introduce greater rigor in our formulations in the future.

It is worth noting, parenthetically, that there is another reason why vagueness enters the behavior theories so far formulated: the neglect of the genuine comparative method in psychology. Terms such as "set," "reinforcement," etc. are being stretched too far if made to refer to *all* organisms, without reference to contextual differences.

In Hull's system we find, in spite of the relatively precise, perhaps overprecise, nature of its formulations, vagueness occurring over such words (undefined terms of the system) as "response." The truth of many of the assertions of Hull's system would appear to depend on the precise interpretation given to this term in various contexts.

"Learning" itself is notoriously a term signifying a process which is vague. Note the sort of distinction made by Humphrey (11). He regards "learning" as a modification of behavior "towards a biologically useful end," but not modification of just as precise a nature towards an end that is not useful, e.g., a seagull learning to follow ships to get food, as opposed to a seagull learning always to fly directly away from ships and starving. This is surely an odd insistence and implies a fundamental confusion, expressible in the statement that learning is not so-and-so, but such-and-such (11, p. 105), which fails to lay emphasis on the essential point that we may define the word learning in any manner we please as long as we have common agreement about that usage. It is unnecessary to add that many different usages of "learning" do in fact exist, although their differences are not always, by any means, made explicit.

The last example we can briefly remark on is the distinction made by some psychologists between insight and trial-and-error types of learning. If learning is not equatable with performance, then it is not immediately obvious that sudden changes of performance are sufficient to build a terminological distinction, and in any event, such distinction needs to be formulated with precision to be useful to science, whereas in the past it appears merely to

refer to different philosophical directives. If this is a genuine nonverbal distinction, then it needs to be formulated in terms that allow some clear operational distinction. This appears typical of the lack of rigor in the existing state of much psychological theory.

Molar and Molecular

A word on the interpretation given to the words "molar" and "molecular" seems apposite as these words are indeed many meaninged.

two points on a continuum, at least in one sense. This distinction is important because propositions which are vague, nontestable, etc. on molar levels may be testable on molecular, and vice versa. This even applies to different molar and different molecular levels. The problem can be simply illustrated diagrammatically. Figure 1 is a representation of the essence of behavior where *S*s stand for stimuli, *R*s for responses and *O* for the organism—to be regarded essentially, of course, as in

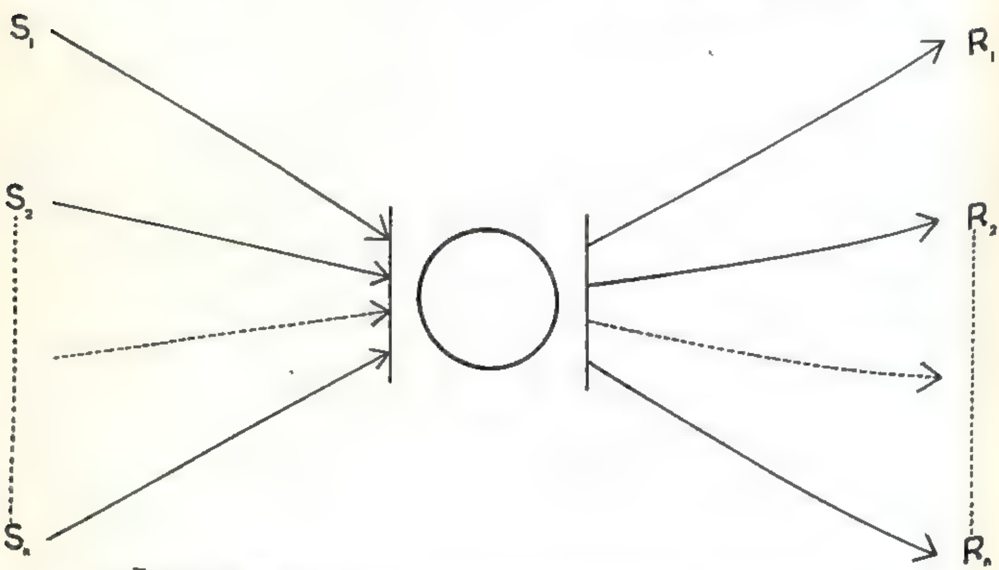


Fig. 1. A schematic representation of a pictorial model of behavior.

Granted that we see all of science on a simple plan of assumptions, consequences of assumptions, tests of consequences, etc.—going on indefinitely in a spiral, where we might equally well have written postulates, theorems, experiments to test theorems, postulates, etc. Granted also that we accept the need for precise propositions on some sort of a hierarchical model as has been suggested, with the added flexibility allowed by glossary and contextual definitions, then we should also expect to find that the molar-molecular apparent dichotomy really represents

an environment and in a space-time manifold. The molar theorist now acts as shown in Fig. 2. He attempts to describe behavior in terms of varied and complex stimulus and response conditions without carrying out tests, observations, etc. on the internal state of the organism. Thus the vagaries that attend making inferences about learning from performance are accounted for by logical constructs (these are represented by the "tags" in Fig. 2, strung together and involving any number). A consideration of latent learning, delayed reward, etc. from

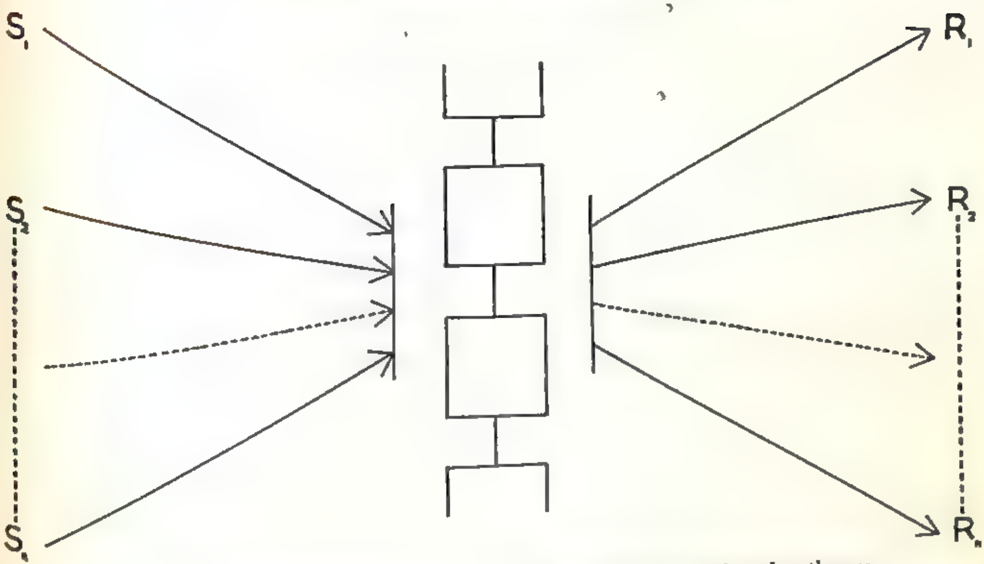


FIG. 2. A schematic representation of a pictorial model of molar theory.

molar theory itself is sufficient to show internal variation cannot always be neglected, even if this was not obvious from our knowledge of physiology, endocrinology, etc. Figure 3 illustrates the molecular viewpoint in essence. Here the molar constructs are broken down into molecular constructs where ⊞ represents diagrammatically the subdivisions of □. The great advance is simply that ⊞ can be regarded as part of □ and part of a molar theory and testable as such, and at the same time is testable on the neurological, endocrinological level. The double purpose is served of making explicit our molar constructs and opening up a further testing ground for their validity.

This whole matter is complex but the term molecular should be regarded

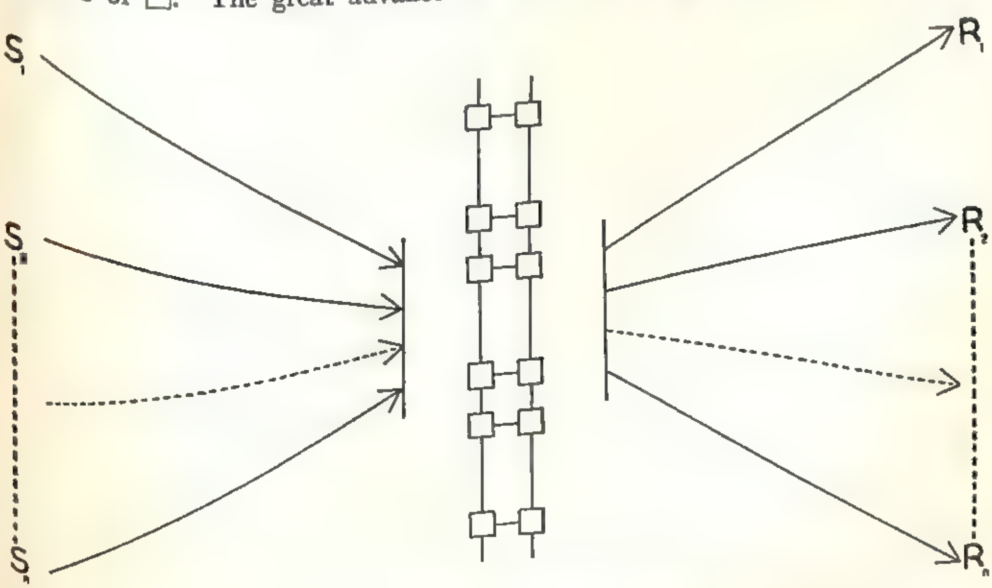


FIG. 3. A schematic representation of a pictorial model of molecular theory.

roughly in the light of tying our molar constructs to the organic data and thus hoping to reduce the work to be done by a single logical construct and thus to reduce the surplus meaning, and also the vagueness.

Summary

It seems necessary to build precise language systems for behavior theory which allow for the breadth of the subject and which at the same time maintain some degree of rigor. The gaps between common languages and narrow postulational techniques need to be filled. The method suggested is straightforward, although sometimes exacting. We must accept a broad postulational form and be prepared to state our undefined terms, postulates, logical constructs, etc. and then allow for a nucleus of undefined terms that are ostensively defined and for "philosophical directives" which should be stated as fully as possible. Then we are left with theorems involving many-meaninged, multiordinal terms and complex undefined concepts, which are best treated by a glossary definition, supplemented by a contextual definition.

Further it will be recognized that we are, in language, on a possible infinity of levels of abstraction and we must differentiate between propositions about psychological theory (this paper is on this level) and propositions within psychological theory.

The attempt has been made to enumerate properties, or in Woodger's terms, to deal with the members of the class as individuals, rather than to talk of classes, as talk of class—more especially definite classes—leads to an unwelcome arbitrariness and vagueness.

Lastly, a few of the more obvious possible verbal confusions have been pointed out in psychological theory as a starting point for a verbal and semantic overhaul which seems vital to the future of behavior theory.

References

1. Ayer, A. J. *Language, truth and logic.* London: Oxford Univer. Press, 1936.
2. Bridgman, P. W. *The logic of modern physics.* New York: Macmillan, 1927.
3. Bridgman, P. W. The nature of some of our physical concepts. I, II, III. *Brit. J. Phil. Sci.*, 1951, **4**, 257–272; **5**, 25–44; **6**, 142–160.
4. Carnap, R. Empiricism, semantics and ontology. *Revue internationale de Philosophie*, 1950, **11**.
5. Feigl, H. Existential hypotheses. *Phil. Sci.*, 1950, **17**, 35–62.
6. George, F. H. Logical constructs and psychological theory. *Psychol. Rev.*, 1953, **60**, 1–6.
7. Gibson, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, 1941, **38**, 781–817.
8. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
9. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
10. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.
11. Humphrey, G. *The nature of learning.* London: Kegan Paul, 1933.
12. Konorski, J. *Conditioned reflexes and neuron organisation.* Cambridge: Cambridge Univer. Press, 1948.
13. Korzybski, A. *Science and sanity.* (2nd Ed.) Lancaster, Pa.: Science Press, 1941.
14. Meehl, P. E. On the circularity of the law of effect. *Psychol. Bull.*, 1950, **47**, 52–75.
15. Symposium on Existential hypotheses. *Phil. Sci.*, 1950, **17**, 164–195.
16. Woodger, J. H. Science without properties. *Brit. J. Phil. Sci.*, 1951, **2**, 193–216.

[MS. received April 3, 1952]

Psychological Review
Vol. 60, No. 4, 1953

CONCEPTS AS OPERATORS[1]

JOE ADAMS

Bryn Mawr College

It is well recognized that ordinarily, if not always, any object that is perceived is perceived either as a certain *kind* of object or as a *familiar* object or both. Almost every one agrees that most examples of this characteristic of perception imply some kind of trace or other prior existing process or organization which has been built up with past experience. Terms which have been used by psychologists to refer to these prior existing somethings are "concepts," "stereotypes," "schemata," "traces," "trace systems," "frames of reference," "enduring cognitive organizations," "cognitive maps," "belief value matrices," "conceptions," "categories," "meanings," etc.[2] A review of the literature employing these terms reveals two interesting points. First, there is not nearly as much overlap as one might expect. Psychologists using the term "concepts" have been interested almost exclusively in concept formation or development. Those using the terms "traces" or "trace systems" (14, 15, 16) have been concerned with the process of communication of the present stimulus with the trace or trace system (the Höffding function) or the communication of one trace with another. Psychologists using the terms "stereotypes," "schemata," or "frames of reference" (1, 2, 26, 27) have dealt with the functional aspects of these constructs in the perceptions, memories, thoughts, and behaviors of the individual. Possibly Krech and Crutchfield (19) with their discussion of cognitive organizations have gone furthest in the integration of these various aspects, but the research reviewed by Vinacke (33) and Leeper (21) illustrates how little integration of this kind has been achieved among psychologists in general.

A second point revealed by a review of the literature is the paucity of adequate terms and methods for the identification, description, and measurement of these constructs. Certainly concepts, etc. must have many characteristics and must reveal their existence in many different ways, but with a relatively few exceptions, psychologists write as though one NUZ is about like any other NUZ, one stereotype about like any other, and as far as investigations of what are called "concepts" go, methods are confined pretty much to sorting or labeling objects, and to formulating criteria, although in the case of "stereotypes" and "schemata" methods involving selective perception and distortions in memory have been employed.

The lack of adequate descriptive terms and methods may contribute to the striking discrepancies between the exciting hypotheses asserted by those who write "in the large" about cognitive processes or linguistic behaviors

[1] This paper, which is intended as a foundation for a forthcoming series of experimental papers, was written during the spring semester, 1952, while the author was a member of the Language and Symbolism project at the University of Michigan, sponsored by the Rockefeller Foundation. It is in part an outgrowth of study during a Social Science Research Council Fellowship, 1948–49, and in part an extension and revision of the author's unpublished Ph.D. dissertation, Princeton University, 1948. Grateful acknowledgment is made to Dr. Heinz Werner and Dr. Silvan Tomkins, who encouraged and advised the author during his SSRC Fellowship.

[2] Some psychologists prefer to deal with this phenomenon in a very different way and do not use these terms or any close equivalents.

and the actual results obtained by experimental research. Many psychologists have tended to ignore the descriptive and functional problems in favor of the genetic problem, i.e., here as in some other areas of psychology, psychologists have been much more interested in giving theoretical explanations of how something has been learned or derived from needs than in describing that something itself and demonstrating its ways of functioning. The history of psychology should give us a clue here. The study of perception made great strides forward when Katz and the gestalt psychologists emphasized the necessity of a description of perceptions rather than an "explaining away" of perception in terms of association or an artificial analysis of perception into hypothetical components, as the structuralists had done. Similarly, the study of behavior has been greatly aided by careful attention to an adequate description (usually in terms of accomplishment, e.g., pressing a bar, reaching a goal box) instead of an analysis into hypothetical components (e.g., reflexes in the Watsonian sense). Likewise, understanding of language is beginning to profit greatly by a careful description of utterances that actually occur (see, for example, the stimulating book by Fries [8]). It may well be that the psychology of cognitive processes will not advance much further until adequate descriptive methods are available.

The purpose of this paper is to propose a number of distinctions which can fruitfully be made in the description of concepts and which enable the use of a common terminology in the discussion not only of concept formation but also of the functioning of concepts in personality. These distinctions are also intended to imply a close relation between traditional psychological problems and more or less psychological hypotheses which have arisen in related fields, especially philosophy, sociology (the sociology of knowledge), and anthropology and linguistics (especially in the writings of Benjamin Whorf [37, 38]). Some indication will also be given of the procedures which are used in the actual application of the descriptive terms. For lack of space the full theoretical and applied value of these distinctions cannot be argued, but a few illustrations are given of how the terms apply to some current research. The terminology used in the paper is, it seems to the author, compatible with the usage of "concept" by psychologists in the fields of concept formation and development. On the other hand, it has been necessary to introduce a number of distinctions which either have not been made explicitly before by psychologists or, if made, have not been labeled. Several of these distinctions are very similar to distinctions made in logic and semantics,[3] and terms used by writers in these fields, especially Carnap (3), will be borrowed. The letter "c" will be prefixed to borrowed terms, to indicate that the usage, though similar, nevertheless has important differences.[4]

Suppose an individual perceives an object X as a certain kind of object, say A. Then we shall say that this individual has a concept C, and we can symbolize the perceptual event as $C(X)=A$. We shall call C an "operator" at the time it is being used to structure perception.

We shall also use the term "operator" for a concept that is being used in

[3] At least to the author these distinctions are similar. I do not agree with those who maintain that discussions of meaning in philosophy and in psychology have, or should have, nothing in common.

[4] The differences between the present usage and Carnap's usage arise from the fact that the terms in the present paper take into account the individual.

an analogous way in cognitive functions other than perception, in the strict sense of the term, such as what is sometimes called "social perception" (e.g., perceiving that Mr. F. likes Miss G., perceiving that Mrs. H. is insincere, etc.), and also in functions like remembering, thinking, imagining, conceiving. In the following discussion, the words "perception" or "perceiving" will be used for brevity, although the assertions are intended to cover other cognitive functions as well. The discussion assumes, therefore, as a working assumption, that the *same internal organizations that operate in one cognitive function operate in other functions.* This assumption is not unverifiable; it is already supported to some extent by available evidence, but if further research should demonstrate that it does not hold, there will be very little value in the notion of concept.

It should be emphasized that "concept" in the present discussion is not being used as a pure intervening variable (23). It is assumed that there are some kinds of internal organizations or processes which are responsible for categorical aspects of cognitive functions, and that these internal organizations have *properties which we must infer from observation.* Part of what follows will seem curiously empty or tautological if this is not kept in mind.

To refer again to our paradigm $C(X) = A$, the significance of the fact that A is not X, that in perceiving the world we use only a small number of the operators that might be used, and that different people, or even the same person at different times, may use different operators, has been discussed by so many psychologists, philosophers, sociologists, anthropologists, and linguists that no attempt will be made to discuss its significance here.[5] It may be pointed out, however, that the question of whether an individual *has* a concept (e.g., whether he can, under suitable instructions, make certain sortings or classifications and/or give a satisfactory definition) is probably not as important as the question of *under what conditions the concept will operate.*

A considerable amount of psychological research can be grouped under the following two problems:

1. What is the functional significance of individual differences in operators under certain standard conditions? Much of the research in projective techniques deals with this problem.

2. What are the factors (other than personality) making for the operation of one concept rather than another in the same individual? Under this heading would come much of the research on the effect of need, set, structural factors, and practice on perception. For example, Duncker's experiments on "functional fixedness of solution objects" (4) can be thought of as dealing with this problem, as he demonstrated that under certain conditions once an object has been operated on by a certain concept, it is more difficult for a different concept to operate on that particular object. Under other conditions, however, the operation of one concept greatly facilitates the operation of certain other concepts on the same object, as for example in racial or ethnic prejudice, where the perception of an individual as a member of a certain group reduces the number or strength of cues or information necessary to trigger the operation of concepts associated with that group. Prob-

[5] Among those who have remarked about the significance of this or related phenomena are Helmholtz, Wundt, James, Titchener, Piaget, Bartlett, Schachtel, Koffka, MacLeod, Sherif and Cantril, Allport and Postman, Krech and Crutchfield, Plato, Aristotle, Aquinas, Locke, Berkeley, Cassirer, De Laguna, Whitehead, Sapir, Whorf, Lady Welby, Lippmann, and Korzybski.

ably almost any operator increases the probability of operation of some concepts and decreases the probability of operation of others.

In our paradigm, $C(X)=A$, we call A a *category* of the concept C, and X an *argument*. We say then that a *concept operates upon an argument to yield a category*.[6] The concept C may have categories in addition to A. By the statement "B and A are both categories of C" we mean that whatever concept operates on an object X to yield A is the same concept as that operating on some object Y to yield B. There are methods for determining whether A and B are categories of the same C, but it will be easier to describe these once we have introduced a few additional terms.

We distinguish between the *category* A and the *verbal label* which the individual uses for the category or for a member of the category. We call such a label an "expressor" of the category A. A category may have more than one expressor. The expressors of a concept are the expressors of its categories.

A complete description of a concept includes two aspects, *content* and *formal properties*. We shall discuss each of these aspects in detail.

The most obvious way of describing concepts is in terms of the assignment of objects to the concept's various categories. In concept-formation experiments this is the most commonly employed test for determining whether a given concept has been formed; implicitly, therefore, this is the way the concept is described. To explicate this notion precisely we shall borrow some terms from logic and semantics. If $C(X)=A$, then we call X a *"c-denotatum"* of A (cf. Morris [24], Russell, J. S. Mill, etc.) and also of any of A's expressors. It is convenient to speak of the *c-extension with respect to a particular set of objects S*. If an individual selects from a set S all those which he is willing to classify as A's, then the latter subset is the c-extension of A with respect to S. The c-extension of a *concept*, on the other hand, is the *set of c-extensions of the categories of the operator;* it is what the mathematician calls a *partition* of a set of objects.

It should be noted that we have not said that any object that *has* certain properties is necessarily a c-denotatum of a certain expressor; it is at this point that we diverge from semantics into pragmatics. An object S is, in our usage, a c-denotatum of an expressor if and only if the individual whose concept we are studying applies that expressor, whether X actually has certain properties or not.

A description of a concept purely in terms of its extensions with respect to certain sets of objects leaves something out, namely what, in a certain sense, the individual "means" by his expressors.[7] We shall call the properties which the individual perceives[8] (or believes or conceives) the c-denotata as having the "c-intension" of the category (and its expressor). In some cases we can distinguish between the *defining c-intension* and the *incidental c-intension*, i.e., between those properties which *define* the category and those which tend to be *attributed* (perhaps with varying strengths of hypotheses)

[6] The present author likes to call these categories *"values,"* as this usage is in accord with the logical analogue (an *operator* operates upon an *argument* to yield a *value*) and is more appropriate than "category" when the notion of a *nominal* operator, as discussed in this paper, is extended to that of an *ordinal*, *interval*, or *ratio* operator.

[7] The inadequacy of a description in terms of extensions is especially obvious in those cases in which the extension is null (e.g., "unicorn") or has only one member (e.g., "myself").

[8] The word "perceives" is used in the broadest sense, to include effects of cues of which the individual may not be conscious.

to any c-denotatum but are not defining properties.[9] For example, part of the defining c-intension of the expressor "Negro" by a prejudiced individual would probably be the property of having dark skin and part of the incidental c-intension might be the property of being innately stupid. *In most cases, however, a distinction between defining and incidental c-intensions will be very difficult, if not impossible, to make.* Even in scientific research these two parts of a c-intension are not always distinguished. If one reviews the history of the duplicity theory of vision, for example, one looks in vain for careful definitions of "rods" and "cones"; as a consequence of this ambiguity, many of the statements made in the development of the theory (including the general statement of the theory itself) are confusing mixtures of definitions, tautologies, and empirical assertions, and the extent to which these statements are verified by certain observations is left indeterminate. It is well recognized that in some cases in which this distinction *can* be made we categorize an object which clearly satisfies a *defining* c-intension, then assume without conscious inference that the *incidental* c-intension is also satisfied. Work on prejudice and stereotypes has emphasized this mechanism, discussed by psychologists under the term "assimilation" (Allport and Postman [1]), "frame of reference" (Sherif and Cantril [27]), or "incorporation into organized cognitive structures" (Krech and Crutchfield [19]).

We have defined "c-intension of a *category*." The defining and incidental c-intensions of a *concept* are the sets of defining and incidental c-intensions of its categories.

Notions like that of c-intension are strongly rejected by many of those interested in problems of language and meaning. The argument runs thus: There are utterances (verbal behaviors), other behaviors, and objects; there may even be images, feelings, and other implicit responses, but look wherever you will, there are certainly no intensions of any kind, and to drag in some notion of properties being attributed to objects is only to introduce confusion, pseudo-problems, and a lot of fruitless verbiage. This argument does not convince the present author; to him the elimination of notions like c-intensions, though perhaps theoretically possible, greatly handicaps any attempt to deal with meaning and leads to rather cumbersome and peculiar circumlocutions.[10]

[9] The term "defining c-intension of the expressor," as here defined, has a meaning somewhat similar to that of Frege's *"Sinn"* (5), Carnap's "intension" (3), Morris's "significatum" (24), Lewis's "signification" (22), J. S. Mill's "connotation," etc., except that these terms are usually used by the authors with reference to languages or semantical systems instead of with reference to one individual's usage, or else they are used in a more restricted sense technically (e.g., cf. Morris). Carnap's "intension" includes the defining c-intension plus all its logical implications, which, in the language of the present paper, may or may not be part of the incidental c-intension. For example, the intension, in Carnap's sense, of the expressor "positive integer" would include the property of having a unique factorization into prime factors, as it has been proved that all positive integers do have unique factorizations, whereas this property would not for any individual's expressor be part of the defining c-intension, and would be a part of the incidental c-intension if and only if the individual using the expressor "positive integer" knew the unique factorization theorem or for some other reason believed that positive integers have this property.

The term "intensional" is used by Korzybski and Hayakawa in a way quite different from the present usage of "defining c-intension," though their usage does have some similarity to that of "incidental c-intension."

[10] One should not assume too hastily that because certain abstract entities do not exist in a spatio-temporal sense they are thereby eliminable in a theoretical analysis. The

There are two methods of investigating c-intensions. Perhaps the most obvious is to ask the subject what he means by a given expressor, thereby obtaining a *formulated c-intension* (a *definition* is a certain kind of formulated c-intension, namely, one which formulates a set of properties both *necessary and sufficient* for class membership; cf. Morris's "formulated significatum" [24]). However, many laboratory investigations as well as everyday observations have shown that formulated c-intensions may be very inadequate. Several investigators (9, 10, 11, 25, 28, 29) have shown this to be the case even in situations where the defining c-intensions can easily be formulated in the language available, which is not the case with certain everyday expressors, such as "dog." Exactly what is meant by "adequacy of a formulated c-intension" will be stated below. At any rate, the obvious inadequacy of formulated c-intensions necessitates the use of the second method, which is *inference from properties which c-denotata have in common,* at least those properties which were *easily perceivable* by the subject during the conditions under which the objects were made c-denotata. It should be emphasized that not all properties held in common by the c-denotata of a category are included in the c-intension of that category; only those which are perceived, believed, or conceived are so included. Further, some properties which the c-denotata do not actually have may be part of the c-intension. When an individual learns that the c-denotata have a property he did not previously know about, that property may become part of the incidental c-intension, though with time it may be absorbed into the defining c-intension.

With the preceding terms we can differentiate several kinds of content similarity between two categories. First, *c-extensional equivalence*[11] with respect to a set of objects *S*, which is simply the degree of extensional overlap (intersection of the two extensions), definable quantitatively. Second, *c-intensional equivalence,* or degree of similarity in c-intensions. In some cases this can be split up into defining c-intensional equivalence[12] and incidental c-intensional equivalence. These terms can also be used with reference to *concepts.* The c-extensional equivalence of two concepts with respect to a set *S* is the overlap[13] in the two respective partitions (sets of c-extensions). The c-intensional equivalence of two concepts is the similarity in the sets of c-intensions of the respective sets of categories.

It was stated earlier that a category may have more than one expressor. Given two expressors, "A" and "B," how do we decide whether they express the same category? There are several criteria to use:

1. Perfect c-extensional equivalence of "A" and "B" with respect to any set of objects.

2. Perfect c-intensional equivalence, in so far as it can be inferred from c-

reader interested in logic and mathematics is referred to Alonzo Church's paper, "The need for abstract entities in semantic analysis," *Proceedings of the American Academy of Arts and Sciences,* Vol. 80, No. 1, July, 1951, for discussion of a related point. For some of the circumlocutions which are necessary to avoid the use of certain abstractions in mathematics, see the paper by Goodman and Quine, "Steps toward a constructive nominalism," *J. symbolic Logic,* 1947, 12, 105–122.

[11] Cf. Carnap's "equivalence" (3), Johnson's "extensional agreement index" (12).

[12] Cf. Carnap's "L-equivalence," which is an either-or kind of predicate not involving similarity.

[13] The overlap of two partitions of a set can be precisely defined in several ways. For the purposes of this paper a definition is not necessary.

extensional equivalence and from formulated c-intensions.

3. A decrease in time required to decide whether to apply expressor "B" if the decision whether to apply expressor "A" has already been made, i.e., a decrease, compared with the time required to make the decision for similar objects not previously decided upon with respect to "A."

4. Indication from the subject, verbal or otherwise, that for objects for which the decision whether to apply "A" has already been made, the decision whether to apply "B" has also been made and explicitly indicated. In other words, that "B" "has the same meaning as 'A.'"

If all these criteria are met, good evidence has been obtained that "A" and "B" express the same category. In some cases criteria 1, 2, and 4 are sufficient, as with some expressors and some sets of objects decision time is so short anyway that no decrease can be observed.

Even if "A" and "B" are not expressors of the same category, they may be expressors of the same concept, i.e., "A" and "B" may be different yet both be categories of "C." It was asserted earlier that there are methods for determining this. We can now list the criteria for deciding whether "A" and "B" are *nonoverlapping categories* of "C":

1. Zero (or near zero) c-extensional equivalence.

2. A decrease in time required to decide whether X is a c-denotatum of "B" when it has already been decided whether X is a c-denotatum of "A."

3. An indication, verbal or otherwise, from the subject that if X has already been made a c-denotatum of "A" then the decision that X is *not* a c-denotatum of "B" has already been made and explicitly indicated.

4. An indication from the subject that he feels or believes that the judgment of whether X is a c-denotatum of "A" is the same kind of judgment as that of whether X is a c-denotatum of "B." (Don't depend on the subject, however, to explicate what he means by "same kind"!)

The fulfillment of these criteria constitutes evidence that "A" and "B" are expressors of *nonoverlapping categories* of "C." The problem of *overlapping* categories or of hierarchical structure is more complicated and no attempt will be made to list criteria here.

Most psychologists would probably agree that almost any word and the same word with "not," or the prefix "un" or "in" preceding it are expressors of the same concept. By using the above criteria we can verify this more or less obvious intuition, at least with human subjects. Certainly any expressors of the form "E" and "Not E" will have zero, or near zero, c-extensional equivalence. This is one way of stating the law of contradiction. It will in general take a subject less time to decide, for certain kinds of objects at least, whether to call an object "E" if he has already decided whether to call it "Not E." Further, if the expressor "E" has already been applied to an object, it will seem to the subject that the same kind of judgment is called for, that he has already decided not to apply the expressor "Not E," and that he has already explicitly indicated that he would not apply "Not E." Words called "opposites" also tend to satisfy these criteria.

Notice that there is no mention above of the following condition: if X is not a c-denotatum of "A" then X *will be* a c-denotatum of "B." This condition will be fulfilled in certain cases, but certainly not in general. Even expressors of the form "E" and "Not E" will not always fulfill this condition, e.g., merely because an individual will not apply the expressor "liberal" to a

given person does not mean that he will apply the expressor "not liberal," and this may hold even if he considers the liberality of the person and tries to make a decision. In other words, in pragmatics a law of excluded middle does not hold.

The foregoing discussion of content description makes it easier to discuss the *formal* properties of concepts. The following is a list of some of these properties:

1. *Structure*. By this is meant simply the number of categories and the relation between them. Possibly there exist concepts with only one category, e.g., concepts expressed by "entity," a word which is applied by some writers to any object, relation, etc., abstract or concrete. Most concepts, however, have at least two categories (though it may be that for many concepts only one category has any denotata), which is the simplest case of *nominal* or qualitative structure. Some nominal structures may be rather complex, as when one category includes two or more other categories. A fairly common hypothesis among psychologists is that, as the individual develops, his concepts have a more and more complex structure (cf. Welch and Welch and Long on hierarchical structure [34, 35]).

Some other types of structure are *ordinal*, *interval*, and *ratio*. It may seem surprising that these terms, developed in the theory of measurement (30), should turn up in this context. Actually, however, the psychology of concepts and the theory of measurement are intimately related, in fact, a really adequate theory of measurement requires an adequate psychology of concepts, whether one calls it that or not, a fact overlooked by some writers in the former field. In the case of interval and ratio operators, and sometimes with ordinal or nominal operators, the operation of looking or otherwise directing the sense organs and making a judgment is supplemented by certain prescribed manipulations of a more or less complex kind, but in all cases the end result is a category or value, and the problem of adequate description in terms of c-extensions, c-intensions, formulations, equivalences, etc., is just as much present in the case of interval and ratio operators (the problem of what the dimension means) as in the case of nominal operators.[14]

It seems apparent that there are important individual differences in operator structure. Frenkel-Brunswik (6, 7) has shown that some children think in dichotomies (weakness-strength, masculine-feminine, right-wrong) more than others, and she maintains that such dichotomous operators are generally characteristic of the ethnically prejudiced and the authoritarian personalities. Klein (13) maintains that a fundamental aspect of personality is revealed in individual differences in operator structure. Korzybski attributed many ills of the world to dichotomous operators ("two-valued, either-or orientations"), and some psychologists more or less agree with this kind of assertion.

2. *Decidability*. We call a concept *decidable* with respect to an argument *X* if *X* can be categorized, in a way satisfactory to the individual, by that concept acting as an operator, otherwise *undecidable*. We call *X* *c-decidable* or *c-undecidable*. The word "vague," when applied to concepts, sometimes means c-undecidability, though other meanings are low c-extensional equivalences and inadequacy of formulated intensions (see below). The inability to tolerate c-undecidability is probably related to intolerance of ambiguity as discussed by Frenkel-Brunswik (6, 7);

[14] Stevens (30) mentions only the most uninteresting case of nominal structure, that in which labels (numerals) are used for identification marks, without meaning (intension).

in her discussion, however, she stresses dichotomous (or other relatively undifferentiated) operators rather than extreme c-undecidability.

3. *Availability.* This is inferred from the conditions under which a concept will or will not operate. A measure of availability can be the amount of information necessary for operation, the speed of operation, etc. Availability is obviously related to the Höffding function.

In an experiment using incomplete figures similar to those of the Street Gestalt Completion Test (31), Verville and Cameron (32) found age and sex differences in the speed of operation of concepts (i.e., time between presentation of a stimulus and the perception of it as a certain kind of thing). Whether this result would hold with other stimulus materials and other conditions of operation has not, to the present writer's knowledge, been investigated, nor has the extent or significance of individual differences in this property been demonstrated. The speed of operation of a concept is analogous to the latency of a response, but this property has not been fully exploited either as a measure of how well a concept has been formed, analogously to the use of latency as a measure of habit strength, or as a diagnostic measure in personality investigations.

4. *Adequacy of formulation of the c-intension.* If an individual formulates in the language L the c-intension of his concept C, then the interpreter I *forms* a concept by *interpreting* the formulated intension. We can distinguish three kinds of adequacy:

a. C-extensional adequacy with expect to a given set S. This is the c-extensional equivalence of the two concepts, C and the interpretation (in L) of the formulated c-intension. Smoke (28) found that some of his subjects' formulated intensions were not completely c-extensionally adequate, *even with respect to the set of objects used as arguments in the formation of the concepts.*

b. Defining c-intensional adequacy. This is the defining c-intensional equivalence of the two concepts.

c. Incidental c-intensional adequacy. This is the incidental c-intensional equivalence of the two concepts.

In scientific arguments over definitions considerable confusion results from failure to specify the kind and degree of adequacy expected or required. This point could with justification be considerably elaborated.

5. *Degree to which the c-intension is conscious.* To some psychologists this property will seem either unintelligible or reducible to property 4. The writer does not share this view; any close tie-up of the psychology of concepts with either psychoanalytic theory or with phenomenology presupposes an investigation of this property.

6. *Validity (with respect to a given set of objects).* The extent to which the c-denotata of the categories *actually satisfy* the respective c-intensions. If one categorizes, as in the Ames demonstration (20) a distorted room as a rectangular room, for this particular object in this situation his operator has very low validity. The validity of operators is one (but not the only) aspect of contact with reality.

7. *Domination.* The extent to which various behaviors and experiences of the individual are functions in a given situation of the result of the operation of the concept. We call such a concept a *dominator* in that situation. Three kinds of domination can be distinguished: *behavior, affect,* and *perception.* Refusing to shake hands with a Negro, feeling negative toward him, and seeing something wrong in everything he does are often examples of behavior,

affect, and perception dominators, respectively.[15]

We can give a partial description of a paranoid condition by saying that the individual is behavior, affect, and perception dominated by highly invalid operators.

In addition to the many problems in the field of concept formation which the foregoing distinctions suggest, we can state, using these distinctions, in fairly precise form a rather comprehensive set of theses about cognitive functioning as follows: Concepts which are describable in terms of formal properties and content[16] play a demonstrable role in the formation and organization of experience and behavior; the formal properties and the content of frequent operators are related to the personality structure of the individual; the content and formal properties of the operators frequently and characteristically used by members of a given culture are strongly related to the lexicon and grammar of the language used in the culture; potential dominators, i.e., concepts which if they operated would dominate, tend to be operators. If any of these theses is true, the psychology of concepts overlaps with many other areas of psychology and has important implications for many other disciplines and applied fields.

[15] It is well recognized that expressors of dominators are sometimes re-defined in an attempt to change the extensions and/or intensions and yet keep the domination. C. S. Stevenson, in his book *Ethics and Language*, calls such a definition a "persuasive definition."

[16] There is a school of thought which holds that concepts, if they are to be talked about at all, must be defined and described in terms of *behaviors* rather than in terms of formal properties and content as described in this paper. It is obvious that behavioral *indices* must be used (such as sorting, discriminatory responses, labeling, autonomic responses, etc.) in the determination of c-extensions and c-intensions; however, the present author believes that any attempt to *define* and *describe* concepts in terms of behaviors will turn out to be about as hopeless and fruitless as was the older attempt to describe concepts in terms of images. Behavior is the result of so many variables that it is difficult to see why anyone regards behavioral *definitions* of any one variable (of any appreciable complexity) as other than a hopelessly complicated task. It should be noted that *descriptions existing in the present literature do describe concepts in terms of c-extensions and c-intensions*, though of course they use behavioral indices in obtaining these descriptions. For a statement of a point of view held by some structural linguists and some psychologists and rather divergent from that presented in this paper, see Bloomfield, L., "Language or ideas?" *Language*, 1936, 12, 89–95.

A more serious objection to the present discussion runs as follows: the author seems to endorse the ancient view that concepts are entities which are carried around by the individual and applied to objects from time to time without taking into account the inner and outer situational contexts, which introduce many changes in these so-called entities. The author readily admits the force of this objection and regards the present treatment as a first approximation only, though one that can lead to important results.

REFERENCES

1. ALLPORT, G., & POSTMAN, L. *The psychology of rumor*. New York: Henry Holt, 1947.
2. BARTLETT, F. C. *Remembering*. Cambridge: Cambridge Univer. Press, 1932.
3. CARNAP, R. *Meaning and necessity*. Chicago: Univer. of Chicago, 1947.
4. DUNCKER, K. On problem solving. *Psychol. Monogr.*, 1945, **58**, No. 5 (Whole No. 270).
5. FREGE, G. Sense and reference. *Philos. Rev.*, 1948, **57**, 209–230.
6. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108–143.
7. FRENKEL-BRUNSWIK, ELSE. Personality theory and perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald, 1951. Pp. 356–419.
8. FRIES, C. C. *The structure of English*. New York: Harcourt, Brace, 1952.
9. HEIDBREDER, EDNA. The attainment of concepts: I. Terminology and methodology. II. The problem. *J. gen. Psychol.*, 1946, **35**, 173–189, 191–223.

10. HEIDBREDER, EDNA. An experimental study of thinking. *Arch. Psychol.*, 1924, No. 73.

11. HEIDBREDER, EDNA. Toward a dynamic psychology of cognition. *Psychol. Rev.*, 1945, **52**, 1–22.

12. JOHNSON, W. *People in quandaries.* New York: Harper, 1946.

13. KLEIN, G. The personal world through perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality.* New York: Ronald, 1951. Pp. 328–355.

14. KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.

15. KOHLER, W. *Gestalt psychology.* New York: Liveright, 1947.

16. KOHLER, W. *Dynamics in psychology.* New York: Liveright, 1940.

17. KORZYBSKI, A. The role of language in the perceptual process. In R. Blake & G. Ramsey (Eds.), *Perception: an approach to personality.* New York: Ronald, 1951. Pp. 170–205.

18. KORZYBSKI, A. *Science and sanity.* (3rd Ed.) Lakeville, Conn.: Int. Non-Aristotelian Library Pub. Co., 1948.

19. KRECH, D., & CRUTCHFIELD, R. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948.

20. LAWRENCE, M. *Studies in human behavior.* Princeton: Princeton Univer. Press, 1949.

21. LEEPER, R. Cognitive processes. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 730–757.

22. LEWIS, C. I. *An analysis of knowledge and valuation.* La Salle, Ill.: Open Court, 1946.

23. MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.

24. MORRIS, C. *Signs, language, and behavior.* New York: Prentice-Hall, 1946.

25. REES, H. J., & ISRAEL, H. E. An investigation of the establishment and operation of mental sets. *Psychol. Monogr.*, 1935, **46**, No. 6 (Whole No. 210).

26. SCHACHTEL, E. On memory and childhood amnesia. *Psychiatry*, 1947, **10**, 1–26.

27. SHERIF, M., & CANTRIL, H. *The psychology of ego-involvements.* New York: Wiley, 1947.

28. SMOKE, K. L. An objective study of concept formation. *Psychol. Monogr.*, 1932, **42**, No. 4 (Whole No. 191).

29. SNYGG, D. The relative difficulty of mechanically equivalent tasks. I. Human learning. *J. genet. Psychol.*, 1935, **47**, 299–320.

30. STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1–49.

31. STREET, R. F. A gestalt completion test: a study of a cross section of intellect. New York: Bur. Publ., Teachers Coll., 1931.

32. VERVILLE, E., & CAMERON, N. Age and sex differences in the perception of incomplete pictures by adults. *J. genet. Psychol.*, 1946, **68**, 149–157.

33. VINACKE, W. E. The investigation of concept formation. *Psychol. Bull.*, 1951, **48**, 1–31.

34. WELCH, L. A preliminary investigation of some aspects of hierarchical development of concepts. *J. genet. Psychol.*, 1940, **56**, 359–378.

35. WELCH, L., & LONG, L. The higher structural phases of concept formation of children. *J. Psychol.*, 1940, **9**, 59–95.

36. WERNER, H. *Comparative psychology of mental development.* (Rev. Ed.) Chicago: Follett, 1948.

37. WHORF, B. An American Indian model of the universe. *Int. J. Amer. Linguistics*, 1950, **16**, 67–72.

38. WHORF, B. *Four articles on metalinguistics.* Foreign Service Institute, Dept. of State, 1949.

[MS. received June 9, 1952]

Psychological Review
Vol. 60, No. 4, 1953

EXPECTANCIES AND HULLIAN THEORY

RICHARD A. BEHAN

Michigan State College

The present paper is written to show that the Hullian theory of behavior is capable of including an expectancy concept which may be likened to that of Tolman (5). It is proposed to derive this concept of expectancy as a theorem to Hull's 1943 postulate set (3). Then, in the language that Hull has preferred, the notion of expectancy will be a corollary. It will be necessary to explain first what is meant by the word *expectancy* in the present paper, and how the notion will be conceptualized. Then the assumptions which are necessary will be set forth, followed by the deduction.

The starting point for the present notion will be the *sign-gestalt-expectation* of Tolman (5). We quote Tolman: "In the case of a sign-gestalt-expectation the organism is ready either as a result of perception (q.v.), mnemonization (q.v.), or inference (q.v.) to have some sort of positive or negative commerce with an immediately presented object (i.e., the sign-object) as the means-end-route (i.e., the signified means-end-relation) to get to or from such and such a further object (i.e., the signified-object)" (5, p. 453). It is thus seen that the sign-gestalt-expectancy is a relation between signs such that the first occurring sign elicits an approach to or an avoidance of the later occurring sign. Special emphasis is placed on the wholeness or the unity of this trinity, and the fact that it directs behavior.

If the above quotation were to be translated into the language of Hullian theory (it is hoped, without doing violence either to Hullian concepts, or to Tolman's terminology), it would read like the following: The goal object evokes a goal response. On successive trials stimuli in the immediate environment come to signal the occurrence of the goal object. Gradually, as the number of trials increases, stimuli which are more distant from the goal object acquire this property of signaling the occurrence of the goal object. Eventually the entire experimental apparatus will possess directive value for the organism. The directive properties which these stimuli acquire result from their relation with the goal object (incentive stimulus). It is this relationship between the incentive stimulus and other stimuli in the experimental situation that constitutes the expectancy.

The expectancy will be conceived as a power of a relation between relations which hold between stimuli and responses, i.e., the expectancy is seen as a relationship which holds between two relations which take stimuli and responses as arguments.[1] First, the relation *evokes,* which holds between a stimulus complex and a response complex, will be considered to be irreflexive and asymmetrical. Second, the relation *leads to,* which holds between response complexes and stimulus complexes, will be considered to be irreflexive and asymmetrical also. The expectancy will then be understood as a power (*po*) of the relation *relative product* holding between the relation *evokes* and the relation *leads to.*

The notion of the relative product is introduced by Whitehead and Russell in their *Principia Mathematica* in Chapter 34 of Volume I (6). The relative

[1] The arguments of a relation are the terms that are related by that relation.

product is a relation between two relations when there is some term to which the members of the domain of the first relation are related by the first relation, and to which the members of the converse domain of the second relation are related by the second relation. That is to say, if we let "R" and "S" denote relations, and "(R/S)" denote their relative product, then the relative product may be defined as

$$(R/S) =_{Df} \hat{x}\hat{z}\{(\exists y)\cdot xRy\cdot ySz\}$$

The power of a relation is the relation between two terms when one is a finite number of relation steps removed from the other. We quote Woodger who says of the relation: "R_{po} is the relation of being related by some power of R. Thus we have $xR_{po}y$ when y can be reached from x in some finite number of R-steps, i.e., when

$$xRv\cdot vRw\cdot \ \ldots\ \cdot zRy$$

$$\text{or } x(R\cup R^2\cup R^3\cup \ldots)y\text{''}$$

(7, p. 39). R_{po} is defined as follows:

$$R_{po} =_{Df} \hat{x}\hat{y}\{(u)\cdot (\breve{R}``u \subset u.\overleftarrow{R}`x \subset u) \supset y \in u.\}$$

R_{po} will be the relation of x to y if it is the case that, whatever class u may be, if $\breve{R}``u$ are included in u and if the $\overleftarrow{R}`x$ are included in u, then $y \in u$."

We are now ready to start the derivation. It is necessary first to find in Hullian theory the relation *evokes* and the relation *leads to*. The first of these relations is easily found in postulate 11 of *Principles of Behavior* (3). Postulate 11 asserts "The momentary effective reaction potential ($_S\dot{\bar{E}}_R$) must exceed the reaction threshold ($_SL_R$) before a stimulus (S) will evoke a response (R)." This postulate provides the relation *evokes*, although the reader will note below that we have taken only the word statement of the postulate. The symbolic statement given in *Principles of Behavior* is defective and we prefer the sentence of assumption 5 to the one listed by Hull in his book. This postulate 11 is now contained in the B part of the 14th postulate in the new postulate set (4).

The notion of a response leading to a stimulus complex is not so easily found. This notion does not appear explicitly in Hull's theory, although it is introduced informally (and incidentally is assumed by many of the students of Hull—and students of students of Hull). For example, the diagram on page 96 of *Principles of Behavior* implies that response may lead to stimulation. Also in some of his earlier work, see especially (2), Hull assumes that response will lead to stimulation. At this point, then, two courses are open: (*a*) One may proceed on convention and simply assume that the Hullian theory contains the notion of a response's leading to stimulation. Or, (*b*) one may stay within the formal theory and derive this notion by the method of substitution. The latter alternative has been chosen in this case for the reason that it is desirable to remain within the theory in the present derivation, and, despite the wide acceptance of the convention mentioned in alternative (*a*), the acceptance of this convention goes beyond the formal theory. It was felt that it was better not to go beyond the theory in cases where it is desirable to show that the theory (the formal theory) contains the notions with which one wishes to work.

Corollary XII in the chapter on *Inhibition and Effective Reaction Potential* asserts: "Whenever conditioned reactions are evoked, whether reinforced or not, reactive inhibition (I_R) is generated" (p. 290). We shall substitute in this corollary the notion of a stimulus complex for the notion of reactive inhibition, obtaining the following lemma:

Lemma I. Whenever conditioned reactions are evoked, whether reinforced or not, stimulus complexes are generated. We may now define the notion of a response leading to stimulation as equivalent to lemma I.

Following is a list of the assumptions which are necessary for the derivation which follows:

1. The logical type of the symbol which denotes *reactive inhibition* is the same as the logical type of the symbol which denotes *stimulus complex.* This assumption is necessary for the justification of the substitution through which we arrived at lemma I from corollary XII.

2. Assume an experimental situation such that the stimulus complexes may occur in serial order, and assume also a succession of trials in this experimental situation. The serial order may be temporal only, or it may be temporal and spatial.

3. Assume that the effective reaction potential of responses with which we wish to deal is in excess of the reaction threshold in each case.

4. Assume the notions of reinforcement and secondary reinforcement.

5. Assume postulate 11.

6. Assume corollary XII.

7. Assume that if the i^{th} stimulus evokes a response, then this response is the i^{th} response.

8. Assume that if the i^{th} response leads to a stimulus complex, then this is the $i+1$st stimulus complex.

Assumptions (7) and (8) specify properties of the experimental situation.

We may now proceed directly to the derivation. Some of the assumptions will be listed below in symbolic language, although not all will enter directly into the derivation. We let "j," "k," "l," and so forth be variables for number signs,

"s_j" denote the jth occurring stimulus complex in the experimental situation,
"r_j" denote the jth occurring response complex in the experimental situation,
"s_i" denote the incentive stimulus, and the i^{th} occurring stimulus complex,
"s_s" denote the stimulus complex at the beginning of the experimental situation, and the s^{th} occurring stimulus complex,
"$s_j\phi r_j$" denote "the stimulus complex s_j evokes the response r_j,"
"$r_j\psi s_k$" denote "the response r_j leads to the stimulus complex s_k."

The symbolic statements of the necessary assumptions follow:

Assumption 3.	$(j, k)s_jEr_k > s_jLr_k$
Assumption 5.	$s_j\dot{\bar{E}}r_j > s_jLr_j . \supset . s_j\phi r_j$
Assumption 6.	$(x):(\exists y).x\epsilon R.y\epsilon I_R. \supset .Ex \supset Gy$
Assumption 7.	$(j, k).s_j\phi r_k \supset j = k$
Assumption 8.	$(j, k).r_j\psi s_k \supset k = j + 1$

Derivation:

Postulate 11	$s_j\dot{\bar{E}}r_j > s_jLr_j . \supset . s_j\phi r_j$	(1)
Assumption 3. (Spec.)	$s_j\dot{\bar{E}}r_j > s_jLr_j$	(2)
(1).(2).Modus Ponens	$s_j\phi r_j$	(3)
Corollary 12	$(x):(\exists y).x\epsilon R.y\epsilon I_R. \supset .Ex \supset Gy$	(4)
(4). $[S/I_R, z/y, \hat{u}(Gu)/G]$	$(x):(\exists z).x\epsilon R.z\epsilon S. \supset .Ex \supset \hat{u}(Gu)z$	(5)
(5). (Concretion)	$(x):(\exists z).x\epsilon R.z\epsilon S. \supset .Ex \supset Gz$	(6)

Define	$r_j\psi s_i =_{Df} (x):(\exists z).x\epsilon R_j.z\epsilon S_i. \supset .Ex \supset Gz$	(7)
(6).(7)	$r_j\psi s_i$	(8)
(3).(8)	$s_j\phi r_j.r_j\psi s_i$	(9)
(9).[$j/i.k/j$]	$s_k\phi r_k.r_k\psi s_j$	(10)
(9).[$1/j.k/i$]	$s_1\phi r_1.r_1\psi s_k$	(11)
	. . .	
(9).[$s/j.1/i$]	$s_s\phi r_s.r_s\psi s_1$	(12)
(12).(11).(10).(9). (Assumption 2)	$s_s\phi r_s.r_s\psi s_1 \ldots s_1\phi r_1.r_1\psi s_k.$ $s_k\phi r_k.r_k\psi s_j.s_j\phi r_j.r_j\psi s_i$	(13)
Define[2]	$(\phi\|\psi) =_{Df} \hat{s}_m\hat{s}_n\{(\exists r_m):s_m\phi r_m.r_m\psi s_n\}$	(14)
(13).(14)	$s_s(\phi\|\psi)s_1 \ldots s_1(\phi\|\psi)s_k.$ $s_k(\phi\|\psi)s_j.s_j(\phi)\|\psi)s_i$	(15)
(15).	$s_s(\phi\|\psi)_{po}s_i.$	(16)

(16) is a relation between the relation *evokes* and the relation *leads to*, i.e., (16), which is the relative product of these two relations, is the expectancy. It will be noted that (15) is just as good as (16), indeed, it is a good deal more expressive. One may think of the process of expectancy development as a process which results in bringing more stimulus complexes into a particular relation with the goal stimulus (incentive stimulus). The expectancy, as it is developed in the present case, is a relation mediated by response, taking stimulus complexes as arguments.

In the quotation from Tolman in the second paragraph of this article the following phrase appears: "In the case of a sign-gestalt-expectancy the organism is ready either as a result of perception (q.v.), mnemonization (q.v.), or inference (q.v.) to have some sort of commerce . . . ." Each of these three terms (perception, mnemonization, and inference) designates a mode of sign-gestalt-expectation. The present writer takes the liberty of assuming that in any actual situation all three modes are active. It is assumed further that Tolman is here referring to a completed expectancy, such as might be similar to our (16). According to lemma I, it is reasonable to assume that during the development of the expectancy the mode of inference plays a very important part.

At this point, a few remarks about lemma I might well be in order. It is obvious that this lemma depends upon corollary 12. If it should ever become necessary to drop corollary 12 from the theory, then lemma I would be eliminated. In such a case, if it were desirable to keep the present notion of expectancy, it would be necessary to make the statement that responses lead to stimulation a part of the formal theory. Such a statement in the form of a postulate would be of great use to a theory like Hull's. For example, such an assumption would allow the

[2] This is the definition of relative product which appears in *Principia Mathematica* (6). The symbolization has been modified to suit the purposes of the present usage.

derivation as a corollary of the postulate of behavioral oscillation (postulate 10) (3).

The relationship between the notion of expectancy, as here presented, and the rest of Hullian theory must now be considered. Every useful construct must have an operational definition and must be related to other constructs in the theory of which it is a part. The beauty of such a derivation as that given above is that it necessitates no new additions to the list of operational definitions in a theory. The derivation provides the necessary reduction chain to reduce the new concept (1). The reduction basis for expectancy in Hullian theory will be the same as that of ${}_{S}\dot{\bar{E}}_{R}$. The derivation also provides the necessary connections with other constructs in the theory. Because the construct expectancy springs from ${}_{S}\dot{\bar{E}}_{R}$, it will automatically become connected with all of the constructs in Hull's theory which contribute to ${}_{S}\dot{\bar{E}}_{R}$.

The present paper will close with a remark about the importance of the possibility of such a derivation as this paper contains for one of the traditional areas of friction in psychology. Here reference is made to the molecular vs. molar controversy. These two terms refer to relative differences between levels of analysis and not to any real fixed differences between approaches to the subject matter of psychology. Those individuals who begin with a relatively molar analysis of the phenomena of psychology end up with a relatively molar analysis of the phenomena they study. Those individuals who begin with a relatively molecular analysis of the phenomena of psychology may, if they utilize the proper techniques, also have a relatively molar analysis of the phenomena they study. The molecularites can have the molar cake and eat it too.

REFERENCES

1. CARNAP, R. Testability and meaning. *Phil. Sci.*, 1936, 3, 420–471; 1937, 4, 1–40.
2. HULL, C. L. Knowledge and purpose as habit mechanisms. *Psychol. Rev.*, 1930, 37, 511–525.
3. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. HULL, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, 57, 173–180.
5. TOLMAN, E. C. *Purposive behavior in animals and men.* Berkeley: Univer. of Calif. Press, 1932.
6. WHITEHEAD, A. N., & RUSSELL, B. *Principia mathematica.* Vol. 1. Cambridge: Cambridge Univer. Press, 1927.
7. WOODGER, J. H. *The axiomatic method in biology.* Cambridge: Cambridge Univer. Press, 1937.

[MS. received July 12, 1952]

Psychological Review
Vol. 60, No. 4, 1953

STIMULUS SATIATION: AN EXPLANATION OF SPONTANEOUS ALTERNATION AND RELATED PHENOMENA [1]

MURRAY GLANZER

Brooklyn College

In 1925 Tolman (20, p. 290) described a peculiar regularity in the behavior of rats that was later labeled spontaneous alternation.

> A simple T maze was used, and it was arranged so that the animal could get back to the food box in identical fashion whether he chose the left or the right angle of the T. Either route met with success. . . . [There was] a very pronounced tendency toward continuous and regular alternation—left, right, left, right, or right, left, right, left. . . . It appeared, in short, that even where either side was equally "satisfactory" there was in our rats a positive tendency left over toward variation of response . . . a positive tendency in and of itself.

Ten years after Tolman's report, spontaneous alternation received intensive study by Dennis and his coworkers. Although these early investigators did not incorporate their data into a formal theory, Dennis (5) did imply that spontaneous alternation was in some way determined by the stimuli the subject faced. For example, the title of one article was "Spontaneous Alternation in Rats as an Indicator of the Persistence of Stimulus Effects," and, in discussion of his results, he wrote about "a tendency to avoid a specific pathway which has recently been traversed" (5, p. 310). This formulation suggests a theory of spontaneous alternation in terms of the stimulus. The rat alternates from one alley (a set of stimuli) to the other (a different set of stimuli), going to Alley B because it has already experienced Alley A. The theory proposed in this paper develops and systematizes this basic approach.

In direct antithesis to a stimulus-oriented theory is one which interprets spontaneous alternation in terms of the response. In a response theory, the animal is viewed as alternating responses instead of alleys, i.e., making a right turn because it has already made a left turn. Many of the more recent investigations examined below have stemmed from this theoretical position.

A Critique of Response Theories

Explanation of spontaneous alternation in terms of the response has been attempted by several Hull-influenced investigators: Heathers (9) using the concept of performance decrement; and Solomon (19) and Zeaman and House (23) using the concept of reactive inhibition. According to both concepts, the occurrence of a response reduces temporarily the probability of its recurrence. The concepts were employed to explain both the fact of spontaneous alternation and the fact, predicted and demonstrated by Heathers (9), that spontaneous alternation disappears as the interval between trials increases.

The concept of reactive inhibition, in addition, assigns an explicit role to the number of times the response is made and the amount of work required by the response. Thus, Zeaman and House (23) predicted and found that giving animals forced trials to one of two

[1] This paper is based upon a Ph.D. dissertation submitted in February 1952 to the department of psychology at the University of Michigan. The writer is indebted to the members of the committee and to Dr. Edward L. Walker, chairman of the committee, who gave invaluable aid in the preparation and editing of this paper.

alleys increased the tendency to alternate to the other; Solomon (19) predicted and found some positive but not conclusive evidence that increased work resulted in increased spontaneous alternation.

Although these experiments seem to give considerable support to an explanation in terms of the response, a number of facts remain that resist explanation by this type of theory. Typical of these are the following:

1. Dennis (5) found that rats did not show spontaneous alternation from unit to unit of a multiple-unit maze. That is, they did not go right-left-right-left within a single trial on a four-unit maze, but showed instead spontaneous alternation in their successive choices at a given choice point. They might go right-right-right-left on the first trial and left-left-left-right on the second trial. This finding is in direct contradiction to the prediction which should follow from a theory in terms of the response.

2. Jackson (11) found that the difference between the two possible responses at the choice point did *not* affect the amount of spontaneous alternation. A Y maze with a small angle of separation, requiring a choice between two practically identical responses, elicited as much spontaneous alternation as a Y make with a wide angle of separation. If the response were the key to spontaneous alternation, then differentiating the response should result in more alternation.

3. Zeaman and House (23), using a procedure of ten forced trials followed by a free trial, and Walker,[2] using simply two free trials, found that rats showed spontaneous alternation with intervals greater than an hour between trials. These findings disagree with expectations on the basis of reactive inhibition which is ordinarily assumed to dissipate within a few minutes.[3]

4. Wingfield (21) found that human subjects showed more spontaneous alternation in choice of lights that differed in color than of lights that were the same. Something in addition to consideration of the subject's responses seems to be demanded by this finding.

5. Dennis (5) varied the amount of work for rats running a maze by increasing the length of the final section from zero to 38 ft. but found no change in the amount of spontaneous alternation. It would seem reasonable to expect that the reactive inhibition from the added run of 38 ft. should dwarf the effects of reactive inhibition from the momentary act of turning right or left.

These difficulties suggest that it might be fruitful to turn to the construction of an explanation in terms of the stimulus. As already pointed out, Dennis (5) anticipated this type of explanation. Recently Montgomery (16, 17, 18), on the basis of empirical findings of his own, has questioned the reactive-inhibition explanation of spontaneous alternation and proposes the concept of an exploratory tendency which is reduced by exposure to a pathway of the maze. Berlyne (2) has presented a similar construct, a curiosity drive, designed to explain curiosity in rats. (Working toward a very different goal, that of incorporating perception and attention as "response" in Hullian theory, Berlyne (3) has also sketched the outlines of a theory that displays some points of similarity to the formulation presented below.)

[2] Walker, E. L. Unpublished data.

[3] Reactive inhibition has to be assumed to disappear quickly in order that it meet its other theoretical commitments. (See Zeaman and House [23].)

A THEORY IN TERMS OF THE STIMULUS

We shall assume that, with continued exposure to an *environment*—to the same *stimuli*—the organism becomes less active in that environment. With an eye to several experiments (5, 9, 19, 23), we assign a specific role to time in the disappearance of this boredom-like effect. Since, moreover, the explanation is in terms of the stimulus, it would be expected that "boredom" created in one situation would carry over to other similar environments and that the more similar two environments are, the greater the carryover.

Using these ideas to form the construct of stimulus satiation,[4] we shall follow Hull's procedure (10) and present the basis of the theory in the form of a postulate.

This new postulate may not only enable us to circumvent the difficulties of the reactive-inhibition explanation but will also yield novel predictions. Furthermore, it may serve to interrelate and explain empirical results from a number of apparently unrelated areas.

The postulate reads as follows:

Each moment an organism perceives a stimulus-object or stimulus-objects, A, there develops a quantity of stimulus satiation to A.

i. *The same amount of stimulus satiation develops in each successive moment. The total amount developed is, therefore, an increasing linear function of time.*

ii. *There is loss of part of each quantity of stimulus satiation in each successive moment. The amount of stimulus satiation remaining from each quantity is a decreasing negative exponential function of time.*

iii. *Stimulus satiation developed to A will be generalized to other stimulus-objects B. The amount of generalized stimulus satiation is an inverse function of the discriminability of A and B.*

iv. *The various quantities of stimulus satiation combine additively.*

v. *Stimulus satiation reduces the organism's tendensy to make any response to A.*

Definition 1. Generalization. When a change in behavior toward A results in a similar change in behavior toward B, all other things being equal, generalization has taken place from A to B.

Definition 2. Discriminability. The ease with which a subject can be induced to make a different response to A than to B, all other things being equal, is called the discriminability of A and B.

Two corollaries follow from the postulate:

Corollary I. As long as an organism remains perceiving A, the amount of stimulus satiation it has to A at a given moment (that is, the total amount developed [*i*] *minus the total amount lost* [*ii*]), *is an increasing negative exponential function of time.*

Corollary II. When an organism stops perceiving A, the amount of stimulus satiation it has to A at a given moment is a decreasing negative exponential function of time.

These corollaries are presented in Table 1 and Fig. 1 with arbitrary values assigned to the functions.

Given the characteristics of various situations, the postulate yields a large number of deductions. A number of these for some basic types of situations are formulated and discussed below. For each deduction discussed, we will indicate whether or not it is crucial for the two theories, e.g., one in terms of the stimulus as against one in terms of the response, as well as state the nature of any existing evidence.

DEDUCTIONS FOR APPROACH RESPONSES IN A TWO-ALTERNATIVE SITUATION

Spontaneous Alternation

Let us start with a very simple situation. This situation is one in which *S* has evenly spaced, consecutive trials in

[4] This term is suggested by Karsten's work (12).

TABLE 1

COROLLARIES I AND II ILLUSTRATED BY OBTAINING VALUES FOR STIMULUS SATIATION, $I_n = 100\ e^{-(n-1)}$ (WHERE n IS THE NUMBER OF MOMENTS SINCE THE APPEARANCE OF THE STIMULUS)*

Stimulus Satiation Quantity	Moments							
	Stimulus Present				Stimulus Absent			
	1	2	3	4	5	6	7	8
I_1	100.0	36.8	13.5	5.0	1.8	.7	.3	.1
I_2		100.0	36.8	13.5	5.0	1.8	.7	.3
I_3			100.0	36.8	13.5	5.0	1.8	.7
I_4				100.0	36.8	13.5	5.0	1.8
ΣI_n	100.0	136.8	150.3	155.3	57.1	21.0	7.8	2.9

* Each row gives the history of a single quantity of stimulus satiation as it decreases with time. Each column gives the total amounts of stimulus satiation present during a single moment.

an apparatus that contains two distinct areas, such as a single-unit T or Y maze. A trial starts with *S*'s entrance into the apparatus and ends with its exit from the apparatus.

In order to complete a trial *S* has to enter one of the two areas or alleys. These areas should be initially more or less equally attractive to *S*. During the course of the experiment, furthermore, one area should not be made more or less attractive than the other by introducing reward or punishment. A situation that has these characteristics will be called a simple two-alternative situation.

The alternatives correspond to stimulus-objects A and B of the postulate. The *S* approaches and perceives one of the two on each trial. It will be assumed that *S*, upon entering an alternative, "perceives" it.

The *S*, perceiving one alternative, builds up stimulus satiation to it that

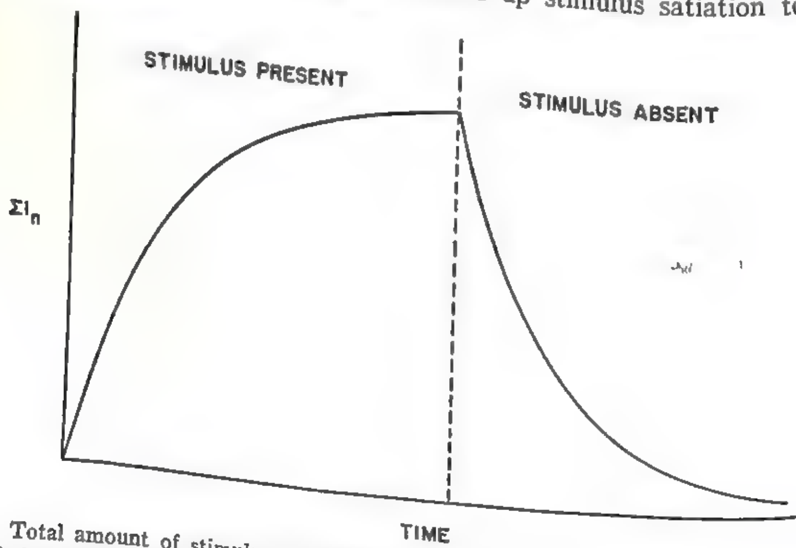


FIG. 1. Total amount of stimulus satiation, ΣI_n, as a function of time and the presence or absence of the stimulus-object. The totals of the columns of Table 1 are represented by this figure.

reduces the tendency to respond to that alternative. The response considered here is the response of approach, since *E* usually records and requires only that response for completion of a trial.

Deduction 1. In the simple two-alternative situation, the subject will alternate choice of arms or alleys in successive trials.

Verified; not crucial.

This is the simple fact of spontaneous alternation which has been repeatedly demonstrated (4, 5, 9, 11, 19, 20, 22, 23).

Deduction 2. In the simple two-alternative situation, if there is an exchange between trials of stimulus-objects so that the cues that were on one side of the *S* now are on the other, and vice versa, *S* will alternate approach to stimulus-objects in successive trials. This will mean a *repetition* of responses rather than an alternation of responses (turns).

Verified; crucial.

This is a key deduction since it contrasts the stimulus satiation and reactive-inhibition explanation most sharply. Stimulus satiation theory predicts a reduction in alternation of response, whereas a theory in terms of the response predicts no decrease.

The author (8) ran rats in a cross-shaped maze with two starting boxes (north and south), making it possible for the *Ss* to start from one starting box for one trial and from the opposite starting box for the next. When this was done, the position of the alleys relative to *S* was reversed: the alley that had been on its right during Trial 1 was on its left for Trial 2. By thus shifting stimuli, it could be determined whether the animals alternated responses as required by the reactive-inhibition explanation or alleys as required by the stimulus satiation explanation.

Twenty-six rats were given two immediately consecutive trials a day for eight days. On four of the days both trials were from the same starting box; on the other four days the trials were from opposite starting boxes. As predicted above, the animals tended to repeat (i.e., go right-right or left-left) rather than alternate their responses on the days that the cues were shifted. This indicates clearly that they were alternating alleys (stimuli) rather than turns (responses). A similar experiment has been described by Montgomery (18) who reports results consistent with these findings.

Deduction 3. In the simple two-alternative situation if an extended series of consecutive trials is given, the tendency toward spontaneous alternation will be at a maximum for the first pair of trials, and will decrease to a minimum in later trials. This decrease will be called cumulation-effect.

Tested but not clearly verified; not crucial.

In a series of consecutive trials in a two-alternative situation, if the first trial was to alternative A, then the second trial would be to alternative B (Deduction 1). There would be two stimulus satiation curves then: one for alternative A starting from the first trial, and continuing its rise through later trials to alternative A; the other for alternative B, starting from the second trial. Since, according to Corollary I, curves of stimulus satiation are negatively accelerated, the two curves would approach each other in the later trials. The difference in amount of stimulus satiation to the two alternatives would therefore decrease as illustrated in Fig. 2, leading to the predicted decrease in spontaneous alternation.

Although results from Wingfield and Dennis (22), Heathers (9) and the author (7) show the general trend toward decrease in spontaneous alternation from first to last trial in a series, the data also show reversals and considerable irregularity.

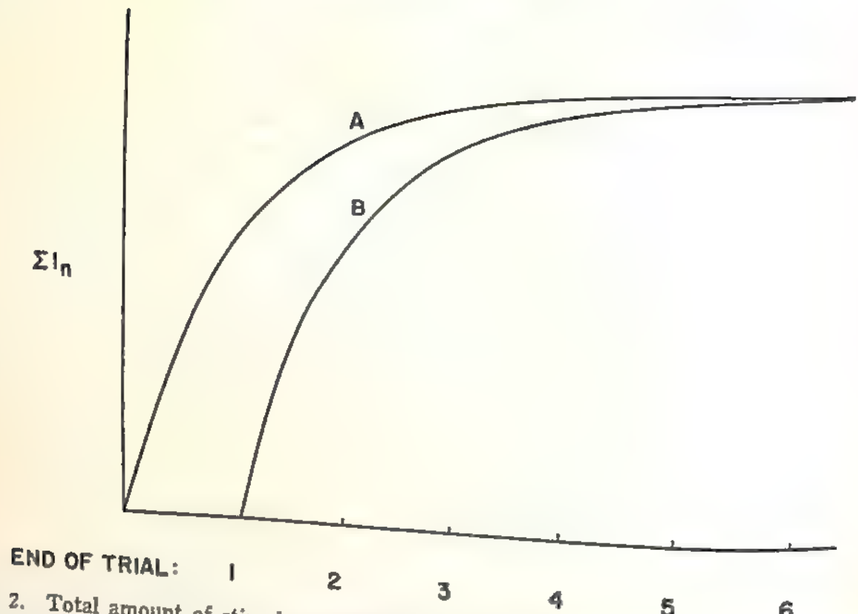


FIG. 2. Total amount of stimulus satiation, ΣI_n, to two alternatives, *A* and *B*, in successive alternating trials. The curves, approaching their limits, cause a decrease in the difference (the distance between the curves) in the amount of stimulus satiation to the two alternatives. All odd-numbered trials are to *A*, even-numbered to *B*. Decreases occurring during intertrial intervals have not been drawn.

The inconclusive nature of these data may result from the very procedure indicated by Deduction 3. In this procedure the animal is free to choose either alternative each trial in the series. The *E* therefore soon loses control of the basic variable, the number and order of particular choices that precede a given trial.

In order to make an adequate test of the prediction, it may be necessary either to use a group of animals large enough to cancel individual differences in the pattern of choices or to change the procedure in order to control the pattern of choices throughout. Clearer results would be expected, then, for the following deduction.

Deduction 4. In the simple two-alternative situation, if a number of forced consecutive alternation trials is given, as the number of trials increases, the tendency toward spontaneous alternation decreases.

Not tested; not crucial.

Now we turn to consideration of the role of generalization. On the basis of part iii of the postulate, the prediction could be made that the more discriminable the alternatives, the greater the amount of spontaneous alternation. Discriminability can be varied by several methods. One is to eliminate or add cues that differentiate the two alternatives.

Deduction 5. In the simple two-alternative situation, as stimuli differentiating the alternatives are eliminated, spontaneous alternation will decrease.

Tested but not verified; crucial.

A study on this by the author (7) comparing spontaneous alternation in Y mazes with many and few differentiating cues yielded negative results. In that experiment, however, there may have been inadequate control of the differentiating cues, since only intramaze cues were varied.

Since discriminability is a function of the organism as well as the environment, the following deductions can also be derived.

Deduction 6. In the simple two-alternative situation, if the ability of the organism to discriminate is reduced, spontaneous alternation will decrease.

Of the conditions of the organism that may affect its ability to discriminate, two may be singled out on the basis of empirical studies. One is sense-organ damage and the other is brain-injury (15). If these conditions may be considered to reduce discriminability, then the following more specific deductions may be made.[5]

Deduction 6*a*. In the simple two-alternative situation, the greater the extent of *S*'s sense-organ damage, the less the amount of spontaneous alternation.

Not tested; crucial.

Deduction 6*b*. In the simple two-alternative situation, the greater the extent of *S*'s brain damage, the less the amount of spontaneous alternation.

Verified; crucial.

Certain of Krechevsky's (13, 14) results support this deduction. He found that brain-operated rats perseverated in choice of one of two paths much more than did normal rats. His situation described below was somewhat complicated but the basic two-alternative arrangement was employed.

Now we turn to the effects of time on spontaneous alternation.

Deduction 7. In the simple two-alternative situation, as the interval between trials increases, the amount of spontaneous alternation decreases.

Verified; not crucial.

This deduction follows simply from part ii of the postulate. It has been verified by Heathers (9) who found that, as the time interval between trials increased from 15 sec. to 15 min., spontaneous altèrnation decreased from 86 to 50 per cent (chance). It is also verified by Zeaman and House (23), who found a decrease over a period of 24 hr.

Deduction 8. In the simple two-alternative situation, the more highly differentiated the alternatives, the longer the time interval between trials necessary for the disappearance of spontaneous alternation.

Not tested; crucial.

This may, in part, explain the widely discrepant data that have been obtained concerning the effects of delay on spontaneous alternation. For example, Walker,[6] using a highly differentiated Y maze, found spontaneous alternation with intervals longer than an hour. Montgomery (16), who took special care to keep his alternatives alike, found that spontaneous alternation disappeared with intervals as short as 90 sec.

Now we turn to an implication of Corollary I, that the longer the time the organism perceives a stimulus-object A, the greater the amount of stimulus satiation to A.

Deduction 9. In the simple two-alternative situation, if *S* is detained in one alternative, the amount of spontaneous alternation to the other alternative on the following trial will be greater, the longer the time of detention.

Not tested; crucial.

The apparent contradiction between Deduction 9 and Deduction 7 may be resolved in the following way. In Deduction 7, the concern is with "interval between trials," an interval spent *outside* the alternative chosen, whereas in Deduction 9, the concern is with a time interval spent within the alternative. These distinctions can be kept clear in

[5] Strictly speaking, Deductions 6*a*, 6*b*, 15 hold only for situations in which there is empirical support for the stated relationship between brain or receptor injury and discriminability.

[6] See footnote 2.

the formulation of a new deduction in terms of the interval between choices.

Deduction 9*a*. In the simple two-alternative situation, with the time interval between choices constant, the greater the proportion of time between choices spent in the presence of the chosen alternative, the greater the amount of spontaneous alternation on the succeeding trial.

Deductions 7 and 9 can now be combined in the following.

Deduction 10. In the simple two-alternative situation, if *S* is delayed between choices and this delay occurs within the alternative last chosen, *S* will show more spontaneous alternation than under a condition of no delay. If, however, *S* is delayed between choices and this delay occurs outside the alternative last chosen, *S* will show less spontaneous alternation than under a condition of no delay.
Verified; crucial.

The author (8) found that when rats were detained in the end box of the alternative chosen in the first run in a T maze, 96 per cent of the group showed spontaneous alternation on the following trial. When they were kept in other parts of the maze at the end of their first run, between 36 and 68 per cent of the group showed spontaneous alternation. When they were not detained, spontaneous alternation varied between 72 and 88 per cent.

Deduction 11. In the simple two-alternative situation, if *S* is given forced runs to one alternative, the greater the number of such forced trials, the greater the amount of spontaneous alternation on a subsequent free-choice trial.
Verified; not crucial.

This procedure is basically the same as that involved in Deduction 9. In Deduction 11 the animal builds up stimulus satiation during repeated visits to one alternative; in Deduction 9 it does so during a period of detention within the alternative. Zeaman and House, deriving their hypothesis from the reactive-inhibition postulate, verified Deduction 11.

The amount of spontaneous alternation obtained by each of the two procedures, by forced trials and by detention at the alternative, may now be compared to yield a clearly critical deduction.

Deduction 12. In the simple two-alternative situation, detention of *S* for a given period of time at the alternative last chosen will yield larger amounts of spontaneous alternation than will a series of forced trials to one alternative over the same period of time.
Not tested; crucial.

In the forced-trial procedure, as compared with the detention procedure, *S*, being taken from end box to starting box, spends a smaller proportion of the time within the alternative itself. Thus *S* has less time to develop stimulus satiation.

Variability (Preference for Variable Situations)

An important variation of the two-alternative situation is one in which the cues of one alternative are varied while the cues of the other are kept constant. This type of situation has been investigated in two studies of "variability." Krechevsky (13, 14) used a two-alternative maze, both arms of which contained a number of turns. One arm could be varied in pattern of turns from trial to trial. Both variable and constant arms were constructed so that there was an equal number of right and left turns within each arm.

Three different maze arrangements were used: a long varying versus a short constant path (13); a short varying versus a long constant path (14); a long constant versus a short constant path (14). In each of these Krechevsky ran two groups of rats, one group normal and the other brain-operated.

In such a situation, stimulus satiation to the varying alternative will be reduced by its division among the different forms of the varying alternative. Essentially, a situation exists in which the constant alternative A is opposed by a family of alternatives B_1, B_2, B_3, . . . B_n. Alternative A bears the weight of all of the stimulus satiation built up during an exposure to it; the family of alternatives B divides stimulus satiation among its variations.

Corollary III. If an alternative is varied from trial to trial, then less stimulus satiation will be built up to it than if no such variation occurs.

Deduction 13. In the simple two-alternative situation, if one of the alternatives is varied from trial to trial and the other is kept constant, then *S* will prefer the varied alternative.

Verified; crucial.

This is verified by Krechevsky (13, 14). Examination of the results for his normal animals shows that the predicted preference is strong enough to cause *S*s to prefer the varied path even when it is the longer path.[7]

It follows that the division of stimulus satiation among the variations of the changing alternative depends on the degree of discriminability of the variations. As the variations become less and less discriminable, the condition of the constant alternative, i.e., of no change, is approached.

Corollary IV. If an alternative is varied from trial to trial, the more discriminable the variations the less the amount of stimulus satiation that will be present for the alternative.

[7] Krechevsky's findings cannot be explained in terms of reactive inhibition or any other theory in terms of the response since, as pointed out earlier, the mazes were designed so that each arm had the same number of right and left turns within it. The *S*s would have, following each trial, an equal amount of reactive inhibition to right and left turning.

There were indicated earlier two ways to reduce discriminability: reduction of the number of distinguishing cues of the perceived objects and impairment of the sensory or nervous systems of the *S*s. Consideration of these methods leads to the following deductions.

Deduction 14. In the simple two-alternative situation, the preference for a varying alternative (as opposed to a constant alternative) will be greater, the greater the difference between variations.

Not tested; crucial.

Deduction 15. In the simple two-alternative situation, brain-injured *S*s will show less preference for a varying alternative (as opposed to a constant alternative) than will normal *S*s.

Verified; crucial.

In Krechevsky's studies (13, 14) the normal animals preferred the varying alternative more strongly than did the operated animals.

Finally, it can be deduced that the greater the number of variations that share the stimulus satiation, the less the amount present to each.

Deduction 16. In the simple two-alternative situation, the greater the number of variations, the greater the preference for the varying alternative.

Not tested; crucial.

DEDUCTIONS FOR THE MULTIPLE-ALTERNATIVE SITUATION

We may now generalize the deductions to apply to situations with more than two alternatives. For example, let us apply the postulate to the behavior of *S* in the three-alternative situation. The *S*, after choosing one alternative, will build up stimulus satiation to it. Upon returning to the choice point for a second trial, *S* will probably choose a new alternative since responsiveness to the first choice has been reduced. If returned again to the

choice point, S will probably choose the "unsatiated" third alternative.

Deduction 1′. In a simple multiple-alternative situation, S will choose many alternatives in a series of consecutive trials.
Definition 3. Many alternatives mean more alternatives than could reasonably be expected on the basis of chance.

Deduction 1′ is verified by the findings of Wingfield and Dennis (22) on the number of different paths chosen by rats running four trials on a four-alternative maze.

Deduction 1 appears as simply a special case of Deduction 1′. In fact, a set of more general deductions can now be obtained in most cases merely by changing "two-alternative" to "multiple-alternative" and "spontaneous alternation" to "tendency to choose many alternatives." Each of the 16 earlier deductions becomes a special case of one of the new set of 16 deductions for multiple-alternative situations.

Consideration of situations involving more than two alternatives suggests a new variable, the number of alternatives, which can form the basis of several new deductions. One such deduction, of particular interest, concerns the cumulation effect of Deductions 3 and 5 and its counterpart in multiple-alternative situations, i.e., the decrease in the tendency to choose many alternatives in an extended series of trials.

Deduction 17′. If there are two simple multiple-alternative situations, and one of them has a greater number of alternatives than the other, then the cumulation effect will be smaller and will develop more slowly for the situation with a greater number of alternatives.

Cumulation effect depends on the summation of amounts of stimulus satiation from successive visits to the same alternative. In the two-alternative situation, S visits alternative A and then B, then A again, then B again, rapidly building up near-maximum amounts of stimulus satiation to both alternatives. According to Deduction 1′, the S, in a three-alternative situation, visits A then B then C and then A again, B again, and C again. The time between successive visits to the same alternative is longer, thereby delaying the onset of maximum stimulus satiation for any single alternative. It will consequently take more time to reach the period when differences in the amounts of stimulus satiation present for the various alternatives approach zero.

Deductions for "Free" Procedure

Deductions have been drawn thus far only for the procedure in which S makes a choice and then is returned by E to the starting point for the next trial. Another procedure, which we shall call the "free" procedure, is one that allows S to make as many choices as it will for a given period of time. The S makes a choice and then returns itself to the choice point for its next choice. When employed without rewards, this is the standard procedure in the study of "exploratory behavior."[8]

Exploratory behavior may now be derived from the stimulus satiation postulate. The deductions can be further generalized to cover this more complex procedure so that exploratory behavior, like spontaneous alternation, becomes a special case governed by the postulate. The word "choice" replaces the word "trial" in the deductions, since we are no longer concerned with a trial as defined earlier; and the procedure is no longer called "simple."

Most of the deductions have not been tested in the "free" procedure. Those

[8] The relationship between spontaneous alternation and exploratory behavior was pointed out by Tolman (20) who explained the former in terms of the latter. We, however, view both spontaneous alternation and exploratory behavior as manifestations of a single underlying factor.

for which there are relevant data have been verified. For example, there is the following deduction involving cumulation effect.

Deduction 3″. In the multiple-alternative situation, if an extended series of consecutive choices is given, as the number of such choices increases, the tendency to choose many alternatives decreases.

This has been verified by Dennis and Sollenberger (6) and later by Montgomery (17). Both studies found that rats enter fewer and fewer alleys in successive periods of free exploration.

Another deduction for which there are data is the following:

Deduction 17″. If there are two multiple-alternative situations and one of them has a greater number of alternatives than the other, then the cumulation effect will be smaller and develop more slowly for the situation with a greater number of alternatives.

This has also been verified by Dennis and Sollenberger (6) and Montgomery (17). They found that larger mazes elicited more activity and continued to elicit activity over a longer period of time than did small mazes. Larger mazes are considered here to contain a greater number of alternatives.

POSSIBLE APPLICATIONS OF THE THEORY TO HUMAN SUBJECTS

Although most of the work in the areas under consideration has been carried out with rats as *Ss*, there is some evidence obtained from human *Ss* that is congruent with the postulate.

The closest analogue to the two-alternative situation that elicits spontaneous alternation in rats is found in a study by Wingfield (21) who required college students to turn on one of a pair of lights. In a series of four trials the *Ss* not only spontaneously alternated the lights they chose to turn on (Deduction 1) but also showed more spontaneous alternation on the first pair than on the second pair of trials (Deduction 3) and more with lights of different hue than with lights of the same hue (Deduction 5).

Karsten's (12) work on the phenomenon which she calls "satiation" (Sättigung) is of particular interest. She had her *Ss* perform repeatedly such tasks as drawing lines or tapping.[9] This activity produces repeated experience of certain stimulus-objects, e.g., lines drawn, and can therefore be considered a case covered by Deduction 13″. This deduction can be considered to predict Karsten's finding of the *Ss*' resistance to continuing the task as well as their variations from the prescribed activity.[10] The phenomenon of cosatiation (transfer of satiation symptoms to new and similar tasks) could also be predicted from part iii of the postulate.

A further finding by Karsten suggests the relevance of the postulate to the problem of fatigue. Symptoms of satiation, including the *Ss*' inability to move their arms, disappeared when the grouping of the lines being drawn was changed even though the required muscle movements remained the same. This recalls two things: part iii of the postulate and the recent emphasis on factors other than simple muscle states in understanding the phenomenon of fatigue (1).

SUMMARY

Because the reactive-inhibition explanation of spontaneous alternation in rats seems inadequate, a new theory in terms of the effects of continued exposure to stimuli is proposed. This is

[9] Berlyne (3) handles these data in the same way in his expansion of Hull's theory to include perceptual "responses."

[10] The *Ss*' resistance also expressed itself in the form of inattentiveness and "fatigue." All of these symptoms could be predicted on the basis of Corollary I.

presented in the form of a postulate with deductions drawn for various aspects of the simple two-alternative situation and available evidence for their validity presented. Following this, the theory is generalized to cover situations with any number of alternatives, and to situations that require other procedures, including one that elicits exploratory behavior. Finally the pertinence of the theory to some aspects of human behavior is indicated. The postulate of stimulus satiation yields novel predictions and serves to unify under a single explanatory construct such diverse phenomena as spontaneous alternation, exploratory behavior, Krechevsky's "variability," and Karsten's "satiation."

REFERENCES

1. BARTLEY, S. H. Fatigue and efficiency. In H. Helson (Ed.), *Theoretical foundations of psychology*. New York: Van Nostrand, 1951. Pp. 318–348.
2. BERLYNE, D. E. Novelty and curiosity as determinants of exploratory behavior. *Brit. J. Psychol.*, 1950, 41, 68–80.
3. BERLYNE, D. E. Attention, perception and behavior theory. *Psychol. Rev.*, 1951, **58**, 137–146.
4. DENNIS, W. A comparison of the rat's first and second exploration of a maze unit. *Amer. J. Psychol.*, 1935, **47**, 488–490.
5. DENNIS, W. Spontaneous alternation in rats as an indicator of the persistence of stimulus effects. *J. comp. Psychol.*, 1939, **28**, 305–312.
6. DENNIS, W., & SOLLENBERGER, R. J. Negative adaptation in the maze exploration of albino rats. *J. comp. Psychol.*, 1934, **18**, 197–206.
7. GLANZER, M. Stimulus satiation as an explanation of spontaneous alternation in rats. Unpublished Ph.D. dissertation, Univer. of Michigan, 1952.
8. GLANZER, M. The role of stimulus satiation in spontaneous alternation. *J. exp. Psychol.*, 1953, **45**, 387–393.
9. HEATHERS, G. L. The avoidance of repetition of a maze reaction in the rat as a function of the time interval between trials. *J. Psychol.*, 1940, **10**, 359–380.
10. HULL, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
11. JACKSON, M. M. Reactive tendencies of the white rat in running and jumping situations. *J. comp. Psychol.*, 1941, **31**, 255–262.
12. KARSTEN, A. Untersuchungen zur Handlungs- und Affektpsychologie: V. Psychische "Sattigung." *Psychol. Forsch.*, 1928, **10**, 142–154.
13. KRECHEVSKY, I. Brain mechanisms and variability: II. Variability where no learning is involved. *J. comp. Psychol.*, 1937, **23**, 139–163.
14. KRECHEVSKY, I. Brain mechanisms and variability: III. Limitations of the effect of cortical injury upon variability. *J. comp. Psychol.*, 1937, **23**, 351–364.
15. LASHLEY, K. S. The mechanism of vision. II. The influence of cerebral lesions upon the threshold of discrimination for brightness in the rat. *J. genet. Psychol.*, 1930, **37**, 461–480.
16. MONTGOMERY, K. C. Spontaneous alternation as a function of time between trials and amount of work. *J. exp. Psychol.*, 1951, **42**, 82–93.
17. MONTGOMERY, K. C. Exploratory behavior and its relation to spontaneous alternation in a series of maze exposures. *J. comp. physiol. Psychol.*, 1952, **45**, 50–57.
18. MONTGOMERY, K. C. A test of two explanations of spontaneous alternation. *J. comp. physiol. Psychol.*, 1952, **45**, 287–294.
19. SOLOMON, R. L. The role of effort in the production of several related behavior phenomena. Unpublished Ph.D. dissertation, Brown Univer., 1947.
20. TOLMAN, E. C. Purpose and cognition: The determiners of animal learning. *Psychol. Rev.*, 1925, **32**, 285–297.
21. WINGFIELD, R. C. Some factors influencing spontaneous alternation in human subjects. *J. comp. Psychol.*, 1943, **35**, 237–244.
22. WINGFIELD, R. C., & DENNIS, W. The dependency of the rat's choice of pathways upon the length of the daily trial series. *J. comp. Psychol.*, 1934, **18**, 135–147.
23. ZEAMAN, D., & HOUSE, B. J. The growth and decay of reactive inhibition as measured by alternation behavior. *J. exp. Psychol.*, 1951, **41**, 177–186.

[MS. received July 2, 1952]

Psychological Review
Vol. 60, No. 4, 1953

THE BACKWARD CURVE: A METHOD FOR THE STUDY OF LEARNING

KEITH J. HAYES

Yerkes Laboratories of Primate Biology, Orange Park, Florida

Graphical representation of the temporal course of improvement in performance has long been a popular research tool in the field of learning theory. Such curves have often been presented as evidence for basic mechanisms presumed to be involved in learning. Potentially, this type of analysis is important: If it can be shown that the acquisition of simple associations is practically always a sigmoid (or linear, or discontinuous, etc.) function of training, then one theory may gain considerably more than another in presumptive validity.

Unfortunately, the type of curve commonly used for this purpose indicates the average performance of a group of *S*s at successive stages of practice. These curves are quite irrelevant to basic problems of learning theory, since their forms are determined not only by the forms of their component individual curves, but also by the distributions of the individual curves. The influence of this second factor is illustrated in Fig. 1, where it is seen that sigmoid average curves tend to result from symmetrical distributions of various kinds of individual curves.[1] When the distribution of individual curves has a strong positive skew, the resulting average curve tends toward predominantly negative acceleration.

The only curves which bear directly on basic theoretical problems are those which represent individual instances of learning—single associations acquired by individual subjects. However, it is difficult to obtain individual curves for the simpler learning situations. Complex tasks such as motor skills, long mazes, or difficult syllable lists may give reliable individual curves, but they are not directly applicable to the more basic problems. Puzzle boxes, discriminations, and conditioned responses are learned so quickly that we can rarely get as many as ten reliable samples of performance during the course of training. Group curves are acceptable substitutes only when we can assume that they are representative of their component individual curves, and this assumption is seldom justified.

Failure of an average curve to represent individual learning is due to individual differences among the subjects. If the only differences are in speed of learning, the general form of the learning function can be approximated reasonably well by a Vincent curve. However, we can seldom justify the assumption that individual curves do not also differ in form. Even if homogeneity of form exists, the irregularity of individual curves makes it difficult to demonstrate. In some cases, form may be systematically related to learning speed, and Hilgard (2, p. 293) has suggested that separate Vincent curves should then be drawn for fast, slow, and perhaps intermediate learning—which is

[1] Normally distributed sigmoid individual curves will give a *similar* average curve; but Merrell (3) has shown (in the context of logistic growth curves) that the average curve need not have the *same* form as its components. The average curve may not be expressible as $y = K/1 + e^{a+bx}$, for instance, even though all of the individual curves are. Sidman (5) has made this same point with reference to the negatively accelerated function, $y = M - Me^{-kx}$.

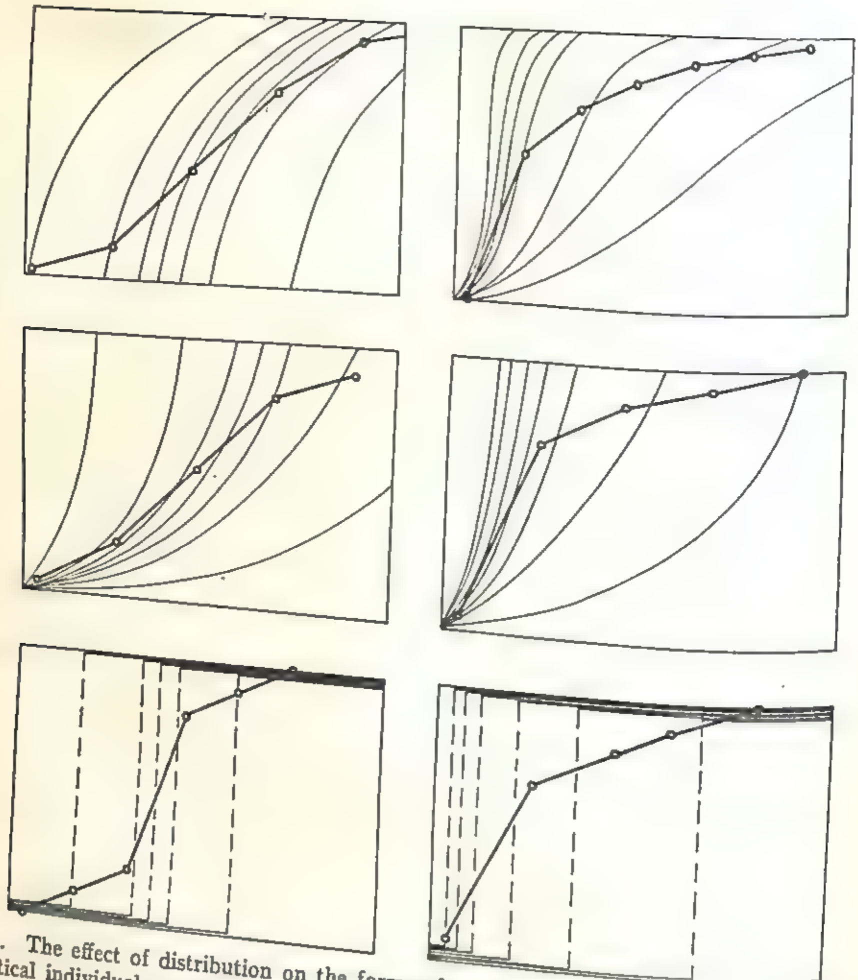


FIG. 1. The effect of distribution on the forms of average learning curves. Light lines show hypothetical individual curves; heavy lines show the resulting average curves. Broken vertical lines indicate discontinuity.

practical only if an adequate number of subjects is available in each of the necessary speed categories.

Where generally heterogeneous groups are involved, it is apparently impossible to plot satisfactory average curves of the complete course of learning. However, certain theoretical issues can be dealt with by examining smaller segments of the learning process, and a method is available for plotting average curves of such segments.[2] The general features of this method may be illustrated by analysis of some data made available by Mr. Robert Thompson.[3] Forty rats were trained in a brightness discrimination problem, to a criterion of ten consecutive correct responses. Motivation was escape from water, and correction was allowed.

The traditional average curve for this experiment is shown in Fig. 2, where each point represents 400 choices—ten trials for each of 40 rats.[4] The light

[2] A procedure very similar to the one described here has been used by Shuttleworth (4) in his study of the prepubertal spurt in human growth.

[3] Formerly with the Yerkes Laboratories, now at the University of Texas.

[4] Means are based on the assumption that no more errors would be made after achieve-

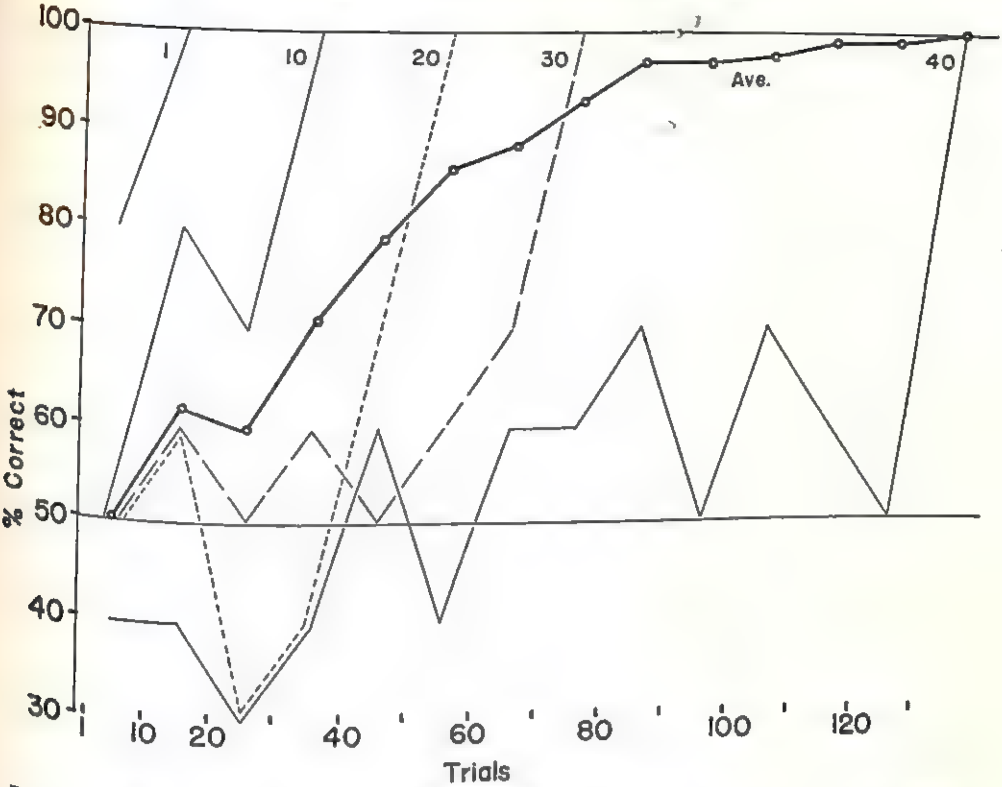


FIG. 2. Brightness discrimination learning in rats. The heavy line is the average learning curve ($N = 40$); light lines are individual curves for the 1st, 10th, 20th, 30th, and 40th rats to reach criterion.

lines in the same figure are individual learning curves for the 1st, 10th, 20th, 30th, and 40th animals to reach criterion. When due allowance has been made for the irregularity of the individual curves, the impression remains that the average curve has grossly distorted the individual data. In contrast to the gradual, negatively accelerated, group curve, the individual curves suggest a variable period of no progress, followed by uniformly sudden (or "insightful") learning.

The form of these individual curves raises the problem of how to plot an average curve which is capable of showing such an abrupt rise to mastery. The proposed solution is shown graphically in Fig. 3. Here the individual curves are displaced horizontally, so that their final points coincide. It is now possible to draw a single curve which does no serious injustice to any of the individuals. Such a curve will not tell us the average accuracy of performance on the x^{th} trial, but it will indicate the average accuracy x trials before achieving the criterion.

The actual construction of this curve involves plotting percentages for successive trials, not working forward from Trial 1, but rather working backward from the criterion. The heavy line in Fig. 4 is such a *backward learning curve,* each point representing the percentage correct on a single trial, for 40 rats.

Backward curves, like Vincent curves, should not be drawn through points based on the criterial trials or the im-

ment of criterion. The "true" curve would be slightly lower, except near the origin.

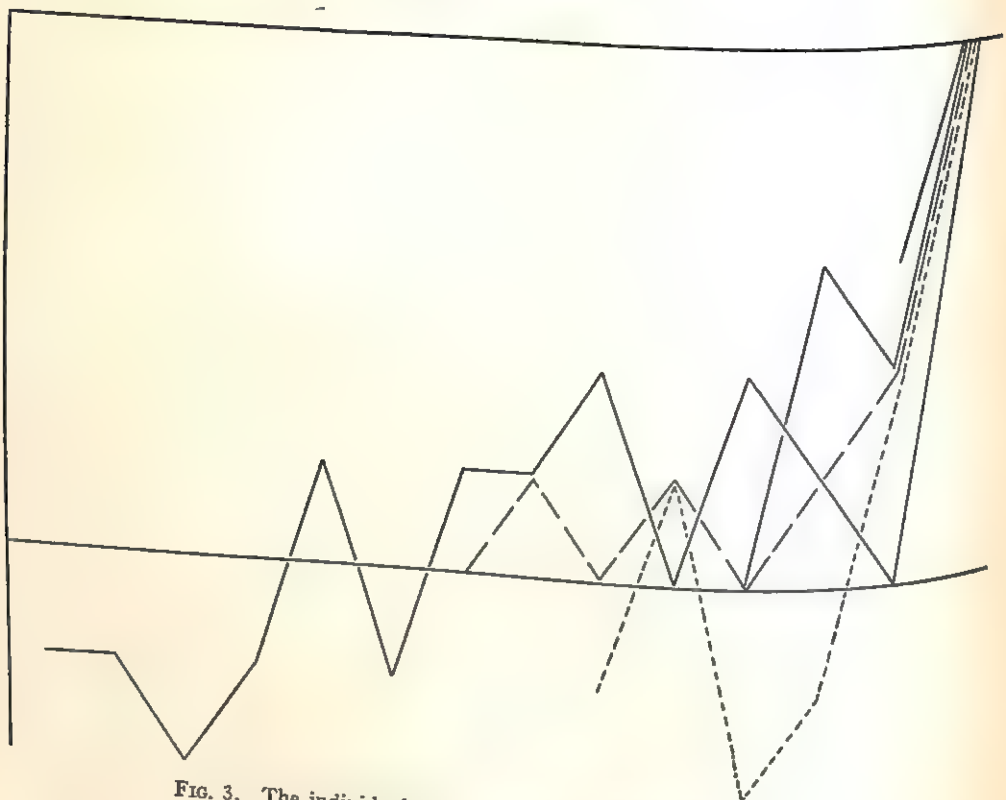


FIG. 3. The individual curves of Fig. 2, shifted horizontally

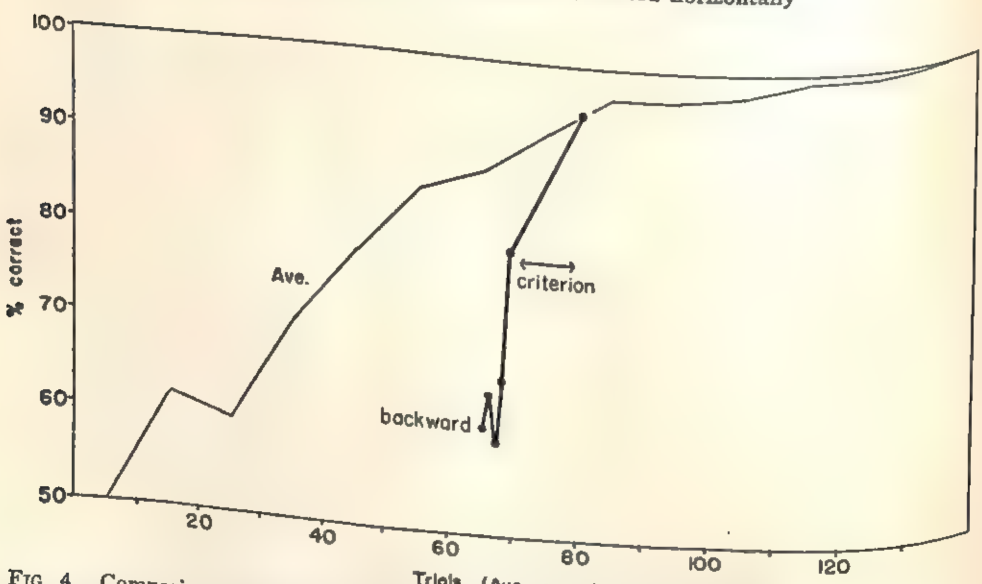


FIG. 4. Comparison of the traditional average curve with the backward curve. The horizontal position of the backward curve was arbitrarily chosen to make the two curves coincide at the 95% level.

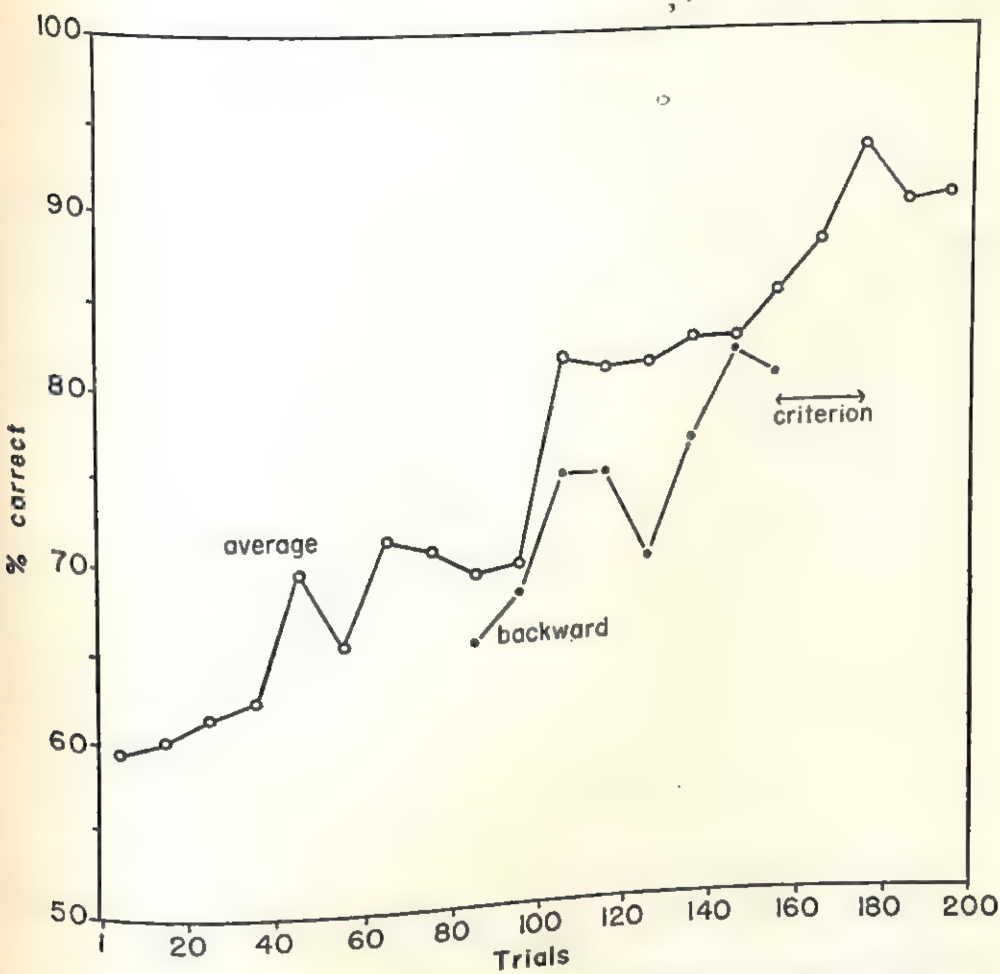


FIG. 5. Average and backward curves for 23 cases of visual discrimination learning by monkeys. Points are based on successive groups of 10 trials. From 90 to 392 trials were required to reach a criterion of 20 successive correct choices. (Data supplied by K. L. Chow.)

mediately preceding trial (2, p. 286).[5] Because of this restriction, it is desirable to have one or more postcriterial trials, to represent the final stage of learning. A postcriterial point was provided in this case by reanalyzing the data in terms of a criterion of nine successive correct trials, instead of the ten originally used in training. (If the rats had been trained with this analysis in mind, they would have been run for several trials beyond criterion, so that postcriterial slope, as well as level, could have been determined.)

[5] Attainment of criterion is due primarily to the subject's increased ability, but it is also due in part to chance fluctuation—which the criterion catches on the upswing. If the criterion is not stringent enough, chance may contribute more than ability. Grant (1) has prepared tables from which appropriate criteria may be selected for discrimination training.

If criterial trials are admitted in a backward curve, "insightful learning" can be demonstrated in a group of tossed coins. This may take time if a stringent criterion is used, but most coins will eventually choose heads ten times in a row if given enough training. Coins differ from other species in showing no retention of the habit on postcriterial trials.

Since some rats learned the problem very quickly, the curve can be carried backward only six trials from the criterion; before this point, progressively fewer rats would be included. Since our purpose is to examine only a small segment of learning, this is not a serious disadvantage. If the curve could be carried further, the phase differences among individual subjects would become progressively larger, and the curve less valid—just as the forward curve becomes less representative as it gets farther from the first trial.

This backward curve is intended to show just one thing—the course of learning in the immediate region of the criterion. Figure 4 indicates that in this experiment performance averaged little better than chance a few trials before the criterion, and that it rose through 80% on the second precriterial trial to 95% just after criterion. The individual curves are too coarse and irregular to give a clear picture of this sudden rise, and the traditional average curve does not show it at all. A Vincent curve for these data would be almost flat, ending below 70%. (A modified Vincent curve, including postcriterial data, would be more appropriate.)

Group data are commonly so heterogeneous that the customary type of average curve is bound to give an impression of gradual learning, regardless of the actual facts. The question may arise as to whether the backward curve is foredoomed to give an impression of sudden learning. The answer is illustrated in Fig. 5, which shows forward and backward curves of equal slope, both based on the same data. The backward curve does demonstrate gradual learning, when it occurs.

Discrimination learning has been used in these examples only because discrimination data happened to be available. The general method is applicable to other learning situations as well and could, in fact, be used more effectively in some other cases. In discrimination work the criterion must include a number of trials which, since they are not used in plotting, leave a large gap in the curve, between pre- and postcriterial points. This gap can be reduced to one trial in conditioning or puzzle-box learning, where the amplitude or time score for a single trial may be used as a criterion.

Although examples have been given of only one application of this method, it has more general possibilities. In addition to detecting sudden solution, it may be used to demonstrate other features (such as plateaus, peaks, or spurts) which would be obscured in the customary over-all group curves. The criteria used in defining the segments to be plotted would not be criteria of mastery, in such cases, and practical difficulties might be encountered in devising criteria which would be reasonably short, yet adequately reliable.

Special analysis was not entirely necessary in the examples given, since many of the subjects learned so slowly that the sudden (or gradual) nature of their final solutions was reasonably apparent in the individual curves. The advantage of the method would be greater in cases of faster learning, where usable individual curves cannot be obtained. It has some utility even with slow learning, however, since large numbers of individual curves are cumbersome to deal with, and impractical to publish.

Speculation as to why the rats of Fig. 4 learned so much more suddenly than the monkeys of Fig. 5 would be hazardous, since the two experiments differed in many ways. It may eventually be profitable to look for the cause of such differences, but only when more cases of sudden learning are available. It is hoped that the procedure sug-

gested here may make some such cases manifest, which might otherwise remain latent in the form of inadequately analyzed data.

SUMMARY

A method is proposed for plotting small segments of group learning curves in such a way that details are shown which would be obscured in the customary types of average curves.

The method is illustrated by search for sudden solution in two batches of discrimination data. Sudden learning was found in one case, and not in the other.

Suggestions are offered for other applications of the method.

REFERENCES

1. GRANT, D. A. Additional tables of the probability of "runs" of correct responses in learning and problem solving. *Psychol. Bull.*, 1947, 44, 276–279.
2. HILGARD, E. R. A summary and evaluation of alternative procedures for the construction of Vincent curves. *Psychol. Bull.*, 1938, 35, 282–297.
3. MERRELL, MARGARET. The relationship of individual growth to average growth. *Human Biol.*, 1931, 3, 37–70.
4. SHUTTLEWORTH, F. K. The physical and mental growth of girls and boys age six to nineteen in relation to age at maximum growth. *Monogr. Soc. Res. Child Develpm.*, 1939, 4, No. 3.
5. SIDMAN, M. A note on functional relations obtained from group data. *Psychol. Bull.*, 1952, 49, 263–269.

[MS. received July 21, 1952]

Psychological Review
Vol. 60, No. 4, 1953

A THEORY OF STIMULUS VARIABILITY IN LEARNING[1]

W. K. ESTES AND C. J. BURKE

Indiana University

There are a number of aspects of the stimulating situation in learning experiments that are recognized as important by theorists of otherwise diverse viewpoints but which require explicit representation in a formal model for effective utilization. One may find, for example, in the writings of Skinner, Hull, and Guthrie clear recognition of the statistical character of the stimulus concept. All conceive a stimulating situation as made up of many components which vary more or less independently. From this locus of agreement, strategies diverge. Skinner (17) incorporates the notion of variability into his stimulus-class concept, but makes little use of it in treating data. Hull states the concept of multiple components explicitly (13) but proceeds to write postulates concerning the conditions of learning in terms of single components, leaving a gap between the formal theory and experimentally defined variables. Guthrie (11) gives verbal interpretations of various phenomena, e.g., effects of repetition, in terms of stimulus variability; these interpretations generally appear plausible but they have not gained wide acceptance among investigators of learning, possibly because Guthrie's assumptions have not been formalized in a way that would make them easily used by others. Statistical theories of learning differ from Hull in making stimulus variability a central concept to be used for explanatory purposes rather than treating it as a source of error, and they go beyond Skinner and Guthrie in attempting to construct a formalism that will permit unambiguous statements of assumptions about stimulus variables and rigorous derivation of the consequences of these assumptions.

It has been shown in a previous paper (7) that several quantitative aspects of learning, for example the exponential curve of habit growth regularly obtained in certain conditioning experiments, follow as consequences of statistical assumptions and need not be accounted for by independent postulates. All of the derivations were carried out, however, under the simplifying assumption that all components of a stimulating situation are equally likely to occur on any trial. By removing that restriction, we are now in a position to generalize and extend the theory in several respects. It will be possible to show that regardless of whether assumptions as to the necessary conditions for learning are drawn from contiguity theories or from reinforcement theories, certain characteristics of the learning process are invariant with respect to stimulus properties while other characteristics depend in specific ways upon the nature of the stimulating situation.

The Generalized Set Model: Assumptions and Notation

The exposure of an organism to a stimulating situation determines a set

[1] This paper is based upon a paper reported by the writers at the Boston meetings of the Institute of Mathematical Statistics in December 1951. The writers' thinking along these and related lines has been stimulated and their research has been facilitated by participation in an interuniversity seminar in mathematical models for behavior theory which met at Tufts College during the summer of 1951 and was sponsored by SSRC.

of events referred to collectively as stimulation. These events constitute the data of the various special disciplines concerned with vision, audition, etc. We wish to formulate our model of the stimulus situation so that information from these special disciplines can be fed into the theory, although utilization of that information will depend upon the demands of learning experiments.

For the present we shall make only the following very general assumptions about the stimulating situation: (*a*) The effect of a stimulus situation upon an organism may be regarded as made up of many component events. (*b*) When a situation is repeated on a series of trials, any one of these component stimulus events may occur on some trials and fail to occur on others; as a first approximation, at least, the relative frequencies of the various stimulus events when the same situation (as defined experimentally) occurs on a series of trials, may be represented by independent probabilities. We formulate these assumptions conceptually as follows:

(*a*) With any given organism we associate a set S^* of N^* elements.[2] The N^* elements of S^* are to represent all of the stimulus events that can occur in that organism in any situation whatever with each of these possible events corresponding to an element of the set. (*b*) For any reproducible stimulating situation we assume a distribution of values of the parameter θ; we represent by θ_i the probability that the stimulus event corresponding to the i^{th} element of S^* occurs on any given trial.

[2] In the sequel, various sets will be designated by the letter S, accompanied by appropriate subscripts and superscripts. The letter N, with the same arrangement of subscripts and superscripts, always denotes the size of the set.

It is assumed that any change in the situation (and we shall attempt to deal only with controlled changes corresponding to manipulations of experimental variables) determines a new distribution of values of the θ_i. By repeating the "same" situation, we mean the same as described in physical terms, and we recognize that, strictly speaking, repetition of the same situation refers to an idealized state of affairs which can be approached by increasing experimental control but possibly never completely realized.

It is recognized that some sources of stimulation are internal to the organism. This means that in order to have a reproducible situation in a learning experiment it is necessary to control the maintenance schedule of the organism and also activities immediately preceding the trial. In the present paper we shall not use the term "trial" in a sufficiently extended sense to necessitate including in the θ distribution movement-produced-stimulation arising from the responses occurring on the trial.

We have noted that the behavior on a given trial is assumed to be a function of the stimulus elements which are sampled on that trial. If in a given situation certain elements of S^* have a probability $\theta = 0$ of being sampled, those elements have a negligible effect upon the behavior in that situation. For this reason we often represent a specific situation by means of a reduced set S. An element of S^* is in S if and only if it has a non-zero value of θ in the given situation. These sets are represented in Fig. 1. In this connection, we must note that a probability of zero for a given event does not mean that the event can never occur "accidentally"; this probability has the weaker meaning that the relative frequency of occurrence of the event is zero in the long run. For a

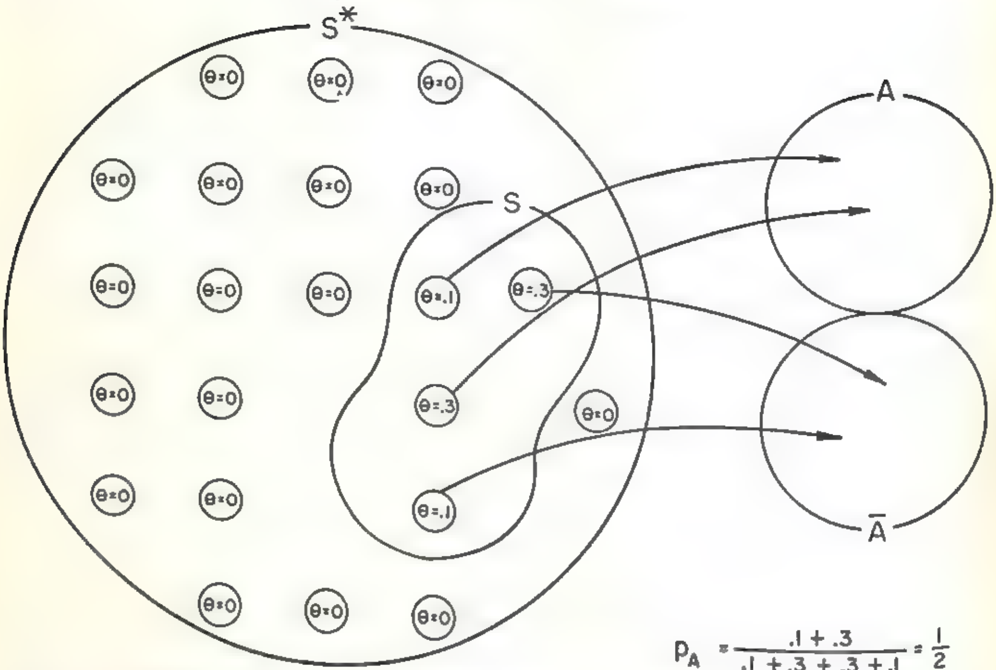


FIG. 1. A schematic representation of stimulus elements, the stimulus space S^*, the reduced set S containing elements with non-zero θ values for a given stimulating situation, and the response classes A and $\bar{A}$. The arrows joining elements of S to the response classes represent the partition of S into S_A and $S_{\bar{A}}$

more detailed explication of this point the reader is referred to Cramér (5).

It should be clearly understood that the probability, θ, that a given stimulus event occurs on a trial may depend upon many different environmental events. For example, a stimulus event associated with visual stimulation may depend for its probability upon several different light sources in the environment. Suppose that for a given stimulus element, the associated probability θ in a given situation depends only upon two separately manipulable components of the environment, a and b, and that the probabilities of the element's being drawn if only a or b alone were present are θ_a and θ_b, respectively. Then the probability attached to this element in the situation with both components present will be

$$\theta = \theta_a + \theta_b - \theta_a\theta_b.$$

THE RESPONSE MODEL

The response model formulated in a previous paper (7) will be used here without any important modification. We shall deal only with the simple case of two mutually exclusive and exhaustive response classes. The response class being recorded in a given situation will be designated A and the complementary class, $\bar{A}$. The dependent variable of the theory is the probability that the response occurring on a given trial is a member of class A. It is recognized that in a learning experiment the behaviors available to the organism may be classified in many different ways, depending upon the interests of the experimenter. The response class selected for investigation may be anything from the simplest reflex to a complex chain of behaviors involving many different groups of effectors. Adequate

treatment of all levels of response specification would require the formulation of a model for the response space and will not be attempted in the present paper. Preliminary investigation of this problem leads us to believe that when a response model is elaborated, the theory developed in this paper will be found to hold rigorously for the most elementary response components and to a first approximation for simple response classes that do not involve chaining. For experimental verification of the present theory we shall look to experiments involving response classes no more complex than flexing a limb, depressing a bar, or moving a key.

CONDITIONAL RELATIONS AND RESPONSE PROBABILITY

We assume that the behavior of an organism on any trial is a function, not of the entire population of possible stimulus events, but only of those stimulus events which occur on that trial; further, when learning takes place, it involves a change in the dependency of the response upon the stimulus events which have occurred on the given trial.

Conditional relations, or for brevity, connections, between response classes and stimulus elements are defined as in other papers on statistical learning theory (3, 7). The response classes A and $\bar{A}$ define a partition of S^* into two subsets $S_A{}^*$ and $S_{\bar{A}}{}^*$. Elements in $S_A{}^*$ are said to be "connected to" or "conditioned to" response A; those in $S_{\bar{A}}{}^*$ to response $\bar{A}$. The concept of a partition implies specifically that every element of S^* must be connected either to A or to $\bar{A}$ but that no element may be connected to both simultaneously.[3] Various features of the model are illustrated in Fig. 1.

For each element in S^* we define a quantity $F_i(n)$ representing the probability that the element in question is connected to response A, i.e., is in the subset $S_A{}^*$, at the end of trial n. The mean value of $F_i(n)$ over S^* is, then, simply the expected proportion of elements connected to A, and if all of the θ_i were equal, it would be natural to define this proportion as the probability, $p(n)$, that response A occurs on trial $n+1$. In the general case, however, not all of the θ_i are equal and the contribution of each element should be weighted by its probability of occurrence, giving

$$(1)\quad p(n) = \frac{\sum_i \theta_i F_i(n)}{\sum_i \theta_i} = \frac{1}{N^*\bar{\theta}} \sum_i \theta_i F_i(n).$$

It will be seen that in the equal θ case, expression (1) reduces to

$$(2)\quad p(n) = \frac{\theta}{N^*\theta} \sum_i F_i(n) = E(F_i(n))$$

which, except for changes in notation, is the definition used in previous papers (6, 7).

The quantity p is, then, another of the principal constructs of the theory. It is referred to as a probability, firstly because we assume explicitly that quantities p are to be manipulated mathematically in accordance with the axioms of probability theory, and secondly because in some situations p can be given a frequency interpretation. In any situation where a sequence of responses can be obtained under conditions of negligible learning and independent trials (as at the asymptote of a simple learning experiment carried out with discrete, well-spaced trials) the numerical value of p is taken as the average relative frequency of response A. For all situations the construct p is assumed to

[3] The argument of this section could as well be given in terms of the set S as of S^*, defining S_A and $S_{\bar{A}}$ as the partition of S imposed by the response classes A and $\bar{A}$.

correspond to a parameter of the behavior system, and we do not cease to speak of this as a probability in the case of a situation where it cannot be evaluated as a relative frequency. It has been shown in a previous paper (7) that p can be related in a simple manner to rate or latency of responding in many situations; thus in all applications of the theory, p is evaluated in accordance with the rules prescribed by the theory, either from frequency data or from other appropriate data, and once evaluated is treated for all mathematical purposes as a probability.

REPRESENTATION OF LEARNING PROCESSES

In order to account for the gradual course of learning in most situations, a number of the earlier quantitative theories, e.g., those of Hull (13), Gulliksen and Wolfle (10), Thurstone (18) have assumed that individual connections are formed gradually over a series of learning trials. Once we adopt a statistical view of the stimulating situation, however, it can be shown rigorously that not only the gradual course of learning but the form of the typical learning curve can be accounted for in terms of probability considerations even if we assume that connections are formed on an all-or-none basis. This being the case, there seems to be no evidence whatsoever that would require a postulate of gradual formation of individual connections. Psychologically an all-or-none assumption has the advantage of enabling us to account readily for the fact that learning is sudden in some situations and gradual in others; mathematically, it has the advantage of great simplicity. For these reasons, recent statistical theories of learning have adopted some form of the all-or-none assumption (3, 7, 15).

Under an all-or-none theory, we must specify the probabilities that any stimulus element that is sampled on a given trial will become connected to A or to $\bar{A}$. For convenience in exposition, we shall limit ourselves in this paper to the simplest special case, i.e., a homogeneous series of discrete trials with probability equal to one that all elements occurring on a trial become connected to response A.

We begin by asking what can be said about the course of learning during a sequence of trials regardless of the distribution of stimulus events. It will be shown that our general assumptions define a family of mathematical operators describing learning during any prescribed sequence of trials, the member of the family applicable in a given situation depending upon the θ distribution. We shall first inquire into the characteristics common to all members of a family, and then into the conditions under which the operators can be approximated adequately by the relatively simple functions that have been found convenient for representing learning data in previous work.

Let us consider the course of learning during a sequence of trials in the simplified situation. Each trial in the series is to begin with the presentation of a certain stimulus complex. This situation defines a distribution of θ over S^* so that each element in S^* has some probability, θ_i, of occurring on any trial, and we represent by S the subset of elements with non-zero θ values; any element that occurs on a trial becomes connected to A (or remains connected to A if it has been drawn on a previous trial). For concreteness the reader might think of a simple conditioning experiment with the CS preceding the US by an optimal interval, and with conditions arranged so that the UR is evoked on each trial and decremental factors are negligible; the situation represented by S is that obtaining from the

onset of the CS to the onset of the US, and the response probability p will refer to the probability of A in this situation. The number of elements in S will be designated by N. For simplicity we shall suppose in the following derivations that none of the elements in S are connected to A at the beginning of the experiment. This means that the learning curves obtained all begin with N_A and p equal to zero. No loss of generality is involved in this simplification; our results may easily be extended to the case of any arbitrary initial condition.

The i^{th} element in S will still remain in $S_{\bar{A}}$ after the n^{th} trial if and only if it is not sampled on any of the first n trials; the likelihood that this occurs is $(1-\theta_i)^n$. Hence, if $F_i(n)$ represents the expected probability that this element is connected to A after the n^{th} trial, we obtain:

$$(3) \qquad F_i(n) = 1 - (1-\theta_i)^n.$$

The expected number of elements in S connected to A after the n^{th} trial, $E[N_A(n)]$, will be the sum of these expected contributions from individual elements:

$$(4) \qquad \begin{aligned} E[N_A(n)] &= \sum_i F_i(n) \\ &= \sum_i [1 - (1-\theta_i)^n] \\ &= N - \sum_i (1-\theta_i)^n. \end{aligned}$$

We are now in a position to express p, the probability of response A, as a function of the number of trials in this situation. By substituting for the term $F_i(n)$ of equation (1) its equivalent from equation (3), we obtain the relation

$$(5) \qquad \begin{aligned} p(n) &= \frac{1}{N\bar{\theta}} \sum_i \theta_i[1-(1-\theta_i)^n] \\ &= 1 - \frac{1}{N\bar{\theta}} \sum_i \theta_i(1-\theta_i)^n. \end{aligned}$$

Equation (5) defines a family of learning curves, one for each possible θ distribution, and it has a number of simple properties that are independent of the distribution of the θ_i. It can easily be verified by substitution that there is a fixed point at $p = 1$, and this will be the asymptote approached by the curve of $p(n)$ vs. n as n increases over all bounds. Members of the family will be monotonically increasing, negatively accelerated curves, approaching a simple negative growth function as the θ_i tend toward equality. If all of the θ_i are equal to $\bar{\theta}$, equation (5) reduces to

$$(6) \qquad p(n) = 1 - (1-\bar{\theta})^n$$

which, except for a change in notation, is the same function derived previously (7) for the equal θ case[4] and corresponds to the linear operator used by Bush and Mosteller (2) for situations where no decremental factor is involved. In mathematical form, equation (6) is the same as Hull's well-known expression for growth of habit strength, but the function does not have the same relation to observed probability of responding in Hull's theory as in the present formulation.

Except where the distribution function of the θ_i either is known, or can be assumed on theoretical grounds to be approximated by some simple expression, equation (5) will not be convenient to work with. In practice we are apt to assume equal θ_i and utilize equation (6) to describe experimental data. The nature of the error of approximation involved in doing this can be stated generally. Immediately after the first trial, the curve for the general case must lie above the curve for the

[4] This is essentially the same function developed for the equal θ case in a previous paper (7); the terms $\bar{\theta}$ and n of equation (6) correspond to the terms $q = s/S$, and T of that paper.

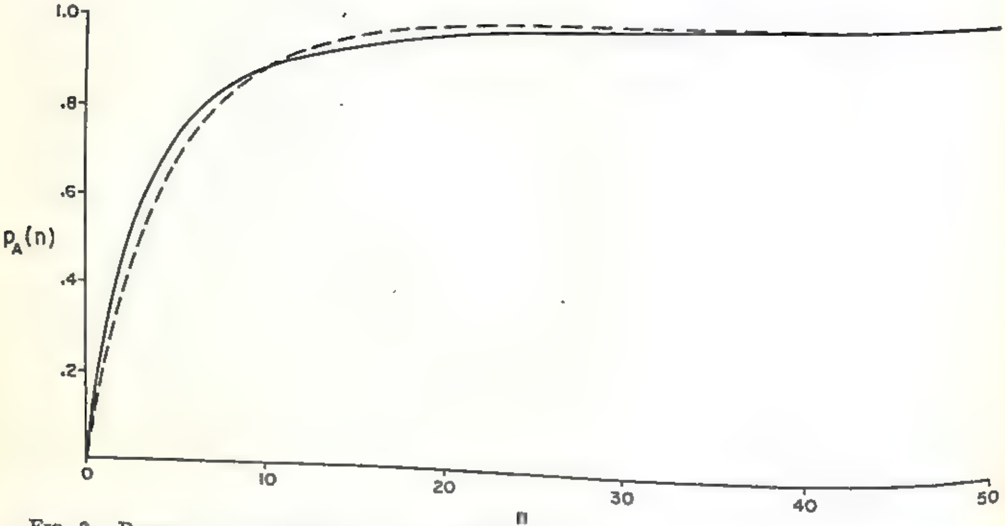


FIG. 2. Response probability, in S, as a function of number of trials for the numerical examples presented in the text. The solid curve is the exact solution for a population of elements, half of which have $\theta = 0.1$ and half $\theta = 0.3$. The dashed curve describes the equal θ approximation with $\bar{\theta} = 0.2$. Initially no elements of S are conditioned to A.

equal θ case; the difference between the two curves increases for a few trials, then decreases until they cross (in constructing hypothetical θ distributions of diverse forms we have usually found this crossing in the neighborhood of the fourth to eighth trial); after crossing, the curves diverge to a smaller extent than before, then come together as both go to the same asymptote at $p = 1$. It can be proved that the curves for the general and special case cross exactly once as n goes from one to infinity. We cannot make any general statement about the maximum error involved in approximating expression (5) with expression (6), but after studying a number of special cases, we are inclined to believe that the error introduced by the approximation will be too small to be readily detectable experimentally for most simple learning situations that do not involve compounding of stimuli.

The development of equations (5) and (6) has necessarily been given in rather general terms, and it may be helpful to illustrate some of the considerations involved by means of a simple numerical example. Imagine that we are dealing with some particular conditioning experiment in which the CS can be represented by a set S, composed of two subsets of stimulus elements, S_1 and S_2, of the sizes $N_1 = N_2 = N/2$, where N is the number of elements in S. Assume that for all elements in S_1 the probability of being drawn on any trial is $\theta_1 = 0.3$ and for those in S_2, $\theta_2 = 0.1$. Now we wish to compute the predicted learning curve during a series of trials on which A responses are reinforced, assuming that we begin with all elements connected to $\bar{A}$. Equation (5) becomes

$$p(n) = 1 - \frac{1}{0.2N}[N_1(0.3) \times (1-0.3)^n + N_2(0.1)(1-0.1)^n]$$
$$= 1 - \frac{1}{0.4}[0.3(0.7)^n + 0.1(0.9)^n].$$

Plotting numerical values computed from this equation, we obtain the solid curve given in Fig. 2.

Now let us approach the same prob-

lem, but supposing this time that we know nothing about the different θ values in the subsets S_1 and S_2 and are given only that $\bar{\theta} = 0.2$. We now obtain predicted learning curves under the equal θ approximation. Equation (6) becomes:

$$p(n) = 1 - (1 - 0.2)^n$$

and numerical values computed from this yield the dashed curve of Fig. 2.

Inspection of Fig. 2 shows that the exact treatment leads to higher values of $p(n)$ on the early trials but to lower values on the later trials, the difference becoming negligible for large n. The reason, in brief, for the steeper curvature of the exact curve is that elements with high θ values are likely to be drawn, and therefore conditioned to A_1 earlier in the learning process than elements with low θ values, and then because they will tend to recur frequently in successive samples, to lead to relatively high values of p. During the late stages of learning, elements with low θ values that have not been drawn on the early trials will contribute more unconnected elements per trial than would be appearing at the same stage with an equal θ distribution and will depress the value of p below the curve for the equal θ approximation.

It should be emphasized that the generality of the present approach to learning theory lies in the concepts introduced and the methods developed for operating with them, not in the particular equations derived. Equation (5), for example, can be expected to apply only to an extremely narrow class of learning experiments. On the other hand, the methods utilized in deriving equation (5) are applicable to a wide variety of situations. For the interest of the experimentally oriented reader, we will indicate briefly a few of the most obvious extensions of the theory developed above, limiting ourselves to the equal θ case.

As written, equation (6) represents the predicted course of conditioning for a single organism with an initial response probability of zero. We can allow for the possibility that an experiment may begin at some value of $p(0)$ other than zero by rewriting (6) in the more general form

$$(7) \quad p(n) = 1 - [1 - p(0)](1 - \bar{\theta})^n$$

which has the same form as (6) except for the initial value.

If we wish to consider the mean course of conditioning in a group of m organisms with like values of $\bar{\theta}$ but varying initial response probabilities, we need simply sum equation (7) over the group and divide by m, obtaining

$$(8) \quad \bar{p}(n) = \frac{1}{m}\sum p(n)$$
$$= 1 - [1 - \bar{p}(0)](1 - \bar{\theta})^n.$$

The standard deviation of $p(n)$ under these circumstances is simply

$$(9) \quad \sigma_p(n) = \sqrt{\frac{1}{m}\sum p^2(n) - \bar{p}^2(n)}$$
$$= (1 - \bar{\theta})^n \sigma_p(0)$$

where $\sigma_p(0)$ is the dispersion of the initial p values for the group. Variability around the mean learning curve decreases to zero in a simple manner as learning progresses.

The treatment of counter-conditioning, i.e., extinguishing one response by giving uniform reinforcement to a competing response, follows automatically from our account of the acquisition process. Returning to equation (6) and recalling that the probabilities of A and $\bar{A}$ must always sum to unity, we note that while response A undergoes conditioning in accordance with (6), response $\bar{A}$ must undergo extinction in

accordance with the function

$$p_{\bar{A}}(n) = 1 - p_A(n) = (1 - \bar{\theta})^n.$$

If, then, we begin with any arbitrary $p(0)$ and arrange conditions so that $\bar{A}$ is evoked and conditioned to all elements drawn on each trial, the extinction of response A will be given by the simple decay function

$$p(n) = p(0)(1 - \bar{\theta})^n. \tag{10}$$

Again the mean and standard deviation of $p(n)$ can easily be computed for a group of organisms with like values of $\bar{\theta}$ but varying values of $p(0)$:

$$\bar{p}(n) = \bar{p}(0)(1 - \bar{\theta})^n \tag{11}$$

$$\sigma_p(n) = (1 - \bar{\theta})^n \sigma_p(0). \tag{12}$$

As in the case of acquisition, variability around the mean curve decreases to zero in a simple manner over a series of trials.

Since variability due to variation in $p(0)$ is reduced during both conditioning and counter-conditioning, it will be seen that in general we should expect less variability around a curve of relearning than around a curve of original learning for a given group of subjects.

Application of the Statistical Model to Learning Experiments

Since our concern in this paper has been with the development of a stimulus model of considerable generality, it has been necessary in the interests of clear exposition to omit reference to most of the empirical material upon which our theoretical assumptions are based. The evaluation of the model must rest upon detailed interpretation of specific experimental situations. It is clear, however, that the statistical model developed here cannot be tested in isolation; only when it is taken together with assumptions as to how learning occurs and with rules of correspondence between terms of the theory and experimental variables, will experimental evaluation be possible. Limitations of space preclude a detailed theoretical analysis of individual learning situations in this paper. In order to indicate how the model will be utilized and to suggest some of its explanatory potentialities we shall conclude with a few general remarks concerning the interpretation of learning phenomena within the theoretical framework we have developed.

Application of the model to any one isolated experiment will always involve an element of circularity, for information about a given θ distribution must be obtained from behavioral data. This circularity disappears as soon as data are available from a number of related experiments. The utility of the concept is expected to lie in the possibility of predicting a variety of facts once the parameters of the θ distribution have been evaluated for a situation. The methodology involved has been illustrated on a small scale by an experiment (6) in which the mean θ value for an operant conditioning situation was estimated from the acquisition curve of a bar-pressing habit and then utilized in predicting the course of acquisition of a second bar-pressing habit by the same animals under slightly modified conditions.

When the statistical model is taken together with an assumption of association by contiguity, we have the essentials of a theory of simple learning. The learning functions (5), (6), and (10) derived above should be expected to provide a description of the course of learning in certain elementary experiments in the areas of conditioning and verbal association. It must be emphasized, however, that these functions alone will not constitute an adequate theory of conditioning, for a number of relevant variables, especially those con-

trolling response decrement, have not been taken into account in our derivations. In conditioning experiments where decremental factors are minimized, there is considerable evidence (1, 4, 9, 14, 16) that the curve of conditioning has the principal properties of our equation (5) and can be well approximated by the equal θ case (7). The fact that functions derived from the model can be fitted to certain empirical curves is a desirable outcome, of course, but cannot be regarded as providing a very exacting test of the theory; probably any contemporary quantitative theory will manage to accomplish this much. On the other hand, the fact that the properties of our learning functions follow from the statistical nature of the stimulating situation is of some interest; in this respect the structure of the present theory is simpler than certain others, e.g., that of Hull (13), which require an independent postulate to account for the form of the conditioning curve.

It should also be noted that deviations from the exponential curve form may be as significant as instances of good fit. From the present model we must predict a specific kind of deviation when the stimulating situation contains elements of widely varying θ values. If, for example, curves of conditioning to two stimuli taken separately yield significantly different values of $\bar{\theta}$, then the curve of conditioning to a compound of the two stimuli should be expected to deviate further than either of the separate curves from a simple growth function. The only relevant experiment we have discovered in the literature is one reported by Miller (16); Miller's results appear to be in line with this analysis, but we would hesitate to regard this aspect of the theory as substantiated until additional relevant data become available.

Although we shall not develop the argument in mathematical detail in the present paper, it may be noted that the statistical association theory yields certain specific predictions concerning the effects of past learning upon the course of learning in a new situation. In general, the increment or decrement in p during any trial depends to a certain extent upon the immediately preceding sequence of trials. Suppose that we have two identical animals each of which has $p(n)$ equal, say, to 0.5 at the end of trial n of an experiment, and suppose that for each animal response A is reinforced on trial $n+1$. The histories of the two animals are presumed to differ in that the first animal has arrived at $p(n)=0.5$ via a sequence of reinforced trials while the second animal has arrived at this value via a sequence of unreinforced trials. On trial $n+1$, the second animal will receive the greater increment to p (except in the equal θ case); the reason is, in brief, that for both animals the stimulus elements most likely to occur on trial $n+1$ are those with high θ values; for the first animal these elements will have occurred frequently during the immediately preceding sequence of trials and thus will tend to be preponderantly connected to A prior to trial $n+1$; in the case of the second animal, the high θ elements will have been connected to $\bar{A}$ during the immediately preceding sequence and thus when A is reinforced on trial $n+1$, the second animal will receive the greater increment in weight of connected elements. From this analysis it follows that, other things equal, a curve of reconditioning will approach its asymptote more rapidly than the curve of original conditioning unless extinction has actually been carried to zero. How important the role of the unequal θ distribution will prove to be in accounting for empirical phenomena of relearning cannot be adequately judged

until further research has provided means for estimating the orders of magnitude of the effects we have mentioned here.

Summary

Earlier statistical treatments of simple associative learning have been refined and generalized by analyzing the stimulus concept in greater detail than heretofore and by taking account of the fact that different components of a stimulating situation may have different probabilities of affecting behavior.

The population of stimulus events corresponding to an independent experimental variable is represented in the statistical model by a mathematical set. The relative frequencies with which various aspects of the stimulus variable affect behavior in a given experiment are represented by set operations and functions.

The statistical model, taken together with an assumption of association by contiguity, provides a limited theory of certain conditioning phenomena. Within this theory it has been possible to distinguish aspects of the learning process that depend upon properties of the stimulating situation from those that do not. Certain general predictions from the theory concerning acquisition, extinction, and relearning, are compared with experimental findings.

Salient characteristics of the model elaborated here are compared with other quantitative formulations of learning.

References

1. Brogden, W. J. Animal studies of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
2. Bush, R. R., & Mosteller, F. A mathematical model for simple learning. *Psychol. Rev.*, 1951, **58**, 313–323.
3. Bush, R. R., & Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.
4. Calvin, J. S. Decremental factors in conditioned-response learning. Unpublished Ph.D. thesis, Yale Univer., 1939.
5. Cramér, H. *Mathematical methods of statistics*. Princeton: Princeton Univer. Press, 1946.
6. Estes, W. K. Effects of competing reactions on the conditioning curve for bar pressing. *J. exp. Psychol.*, 1950, **40**, 200–205.
7. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
8. Feller, W. *An introduction to probability theory and its applications*. New York: Wiley, 1950.
9. Grant, D. A., & Hake, H. W. Dark adaption and the Humphreys random reinforcement phenomenon in human eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 417–423.
10. Gulliksen, H., & Wolfle, D. L. A theory of learning and transfer. *Psychometrika*, 1938, **3**, 127–149.
11. Guthrie, E. R. Psychological facts and psychological theory. *Psychol. Bull.*, 1946, **43**, 1–20.
12. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning*. New York: Appleton-Century, 1940.
13. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
14. Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
15. Miller, G. A., & McGill, W. J. A statistical description of verbal learning. *Psychometrika*, in press.
16. Miller, J. The rate of conditioning of human subjects to single and multiple conditioned stimuli. *J. gen. Psychol.*, 1939, **20**, 399–408.
17. Skinner, B. F. *The behavior of organisms*. New York: Appleton-Century-Crofts, 1938.
18. Thurstone, L. L. The learning function. *J. gen. Psychol.*, 1930, **3**, 469–493.

[MS. received November 12, 1952]

VOL. 60, No. 5 SEPTEMBER, 1953

THE PSYCHOLOGICAL REVIEW

A RE-EXAMINATION OF THE CONCEPT OF INSTINCT

W. C. ALLEE

Department of Biology, University of Florida

HENRY W. NISSEN

Yale University and Yerkes Laboratories of Primate Biology, Inc.

AND MEYER F. NIMKOFF

Department of Sociology, The Florida State University

Introduction

[The idea of inviting experts or authorities in related fields to compare and evaluate their knowledge with regard to certain common concepts, is not new—although it has certainly been tried too infrequently. This plan was adopted at the annual convention of the Florida Psychological Association at Daytona Beach, Florida, in April 1952, when a zoologist, a psychologist, and a sociologist pooled their resources in reevaluating the concept of instinct.

All were leaders in their fields. Dr. W. C. Allee, Head of the Department of Biology at the University of Florida, and formerly Professor of Zoology and Dean at the University of Chicago, presented the zoologist's position. Dr. Henry W. Nissen, Assistant Director of the Yerkes Laboratories of Primate Biology at Orange Park, Florida, and Professor of Psychology at Yale University, spoke for the psychologists. And Dr. Meyer F. Nimkoff, Head of the Department of Sociology at Florida State University, and eminent authority on the family, represented the sociologists. The symposium was under the direction and moderation of Dean Stanley E. Wimberly of the University of Florida, who asked that the speakers present their material in an evolutionary or phylogenetic order, the zoologist coming first, the psychologist second, and the sociologist third.

The contributions of the participants were so effective that the symposium became without question the high light of the two-day convention. Plans for its publication were immediately begun. What follows represents the culmination of these efforts. Since the opportunities for citing references on such a topic are almost limitless, each participant was asked not to present a bibliography.—W. N. Kellogg, President, Florida Psychological Association.]

Instinct from the Zoologist's Standpoint

W. C. Allee

A simple outline of different phases of animal behavior may help orient this discussion of instinct. One such outline recognizes a primary division of all behavior into (I) unlearned, and (II) learned behavior patterns. The former can be broken into smaller categories as follows: (*a*) poorly organized responses of animals lacking nervous systems, e.g., sponges, or early embryos; (*b*) reflex-arc reactions; (*c*) kineses, meaning a speeding or slowing of unoriented responses of whole organisms; (*d*) the oriented tropisms of sessile plants and animals and the tactic reactions of motile forms consisting, when shown in purest form, of oriented forced movements which seem to resemble reflex actions of entire organisms; (*e*) instincts,

the most complex of all phases of unlearned behavior, including many levels of behavior from variable appetitive responses to more rigid consummatory actions.

The great confusion about instincts results in part from the fact that much so-called instinctive action of man and other vertebrates, and, perhaps, of some invertebrates also is really behavior based on very early training. The training has been forgotten; the definite activity pattern remains. Sometimes the learning may be based on a single exposure to a given situation.

There are a great many definitions of instinct. These fall into two main types: (*a*) those that are primarily objective, the biologically behavioristic definitions; (*b*) the subjective ones.

Existence of the subjective emphasis is a part of the cause of justifiable skepticism regarding instincts among biologists, including, of course, psychologists. Parenthetically, the scientific part of psychology obviously is a branch of biology. The subjective definitions may be illustrated by the statement that if birds migrate south early in the fall, earlier than they are normally expected to do, this may be taken as a sign that an uncommonly cold winter is coming. The migrating birds are supposed to have a supernatural instinctive insight in the matter. Such fantasies are similar to folktales about whiteness of chicken bones foretelling the severity of the coming winter.

Good definitions of instinct are hard come by. Most of them, even of the objective, biologically behavioristic sort, are involved and difficult to interpret. There are two I wish to present. One is adapted from W. M. Wheeler, who said, approximately, that an instinct is a relatively complicated activity of an organism that is acting (*a*) as a whole rather than as a part; (*b*) as a representative of the species rather than as an individual; (*c*) without previous experience or without modification caused by experience; and (*d*) with an end or purpose of which the reacting animal has no knowledge. More recently, in 1951, N. Tinbergen, in his book entitled *The Study of Instinct*, gives his definition only after he is more than halfway through the book. Even then, the definition is offered almost apologetically in loose association with his admittedly somewhat imprecise discussion of neurophysiological relationships. His statement is (p. 112): "I will tentatively define an instinct as a hierarchically organized nervous mechanism which is susceptible to certain priming, releasing, and directing impulses, of internal as well as of external origin, and which responds to these impulses by coordinated movements that contribute to the maintenance of the individual and the species." These two objective definitions supplement each other more than they overlap. They may be tested against such activity patterns as breathing, swallowing, gland secretion, muscle contraction, all of which may be shown to occur without learning. Some types of glandular secretion, for example, can be effected by learning only with great difficulty, if at all. On casual inspection, such activities as breathing, etc. come more nearly to fitting Tinbergen's definition than they do that advanced by Wheeler, although they run aground there only on the first idea that the organism should be acting as a whole rather than as a part when displaying instinctive behavior.

Instinctive reactions represent examples of prolongation of development exhibited as behavior rather than as formation of new structure. Contrast, for example, the difference between a chrysalid and a cocoon. Each is found at pupation time in the life history of certain members of the Lepidoptera. The butterflies form a chrysalid which con-

sists essentially of a thickening of the hypodermis to form a protective covering. This thickened chrysalid is frequently attached to the walls of some crevice by a gluelike secretion. Inside its protecting covering the pupation processes take place. In contrast, the moths, as pupation time approaches, spin a more or less elaborate silk cocoon. Inside the cocoon the moth larva also develops a somewhat thickened hypodermis, which serves as an additional protection. There is no essential difference in these activities being examples of prolongation of development. One is limited primarily to the hardening of the hypodermis; the other is complicated by the presence of silk-spinning activities which represent instinctive behavior. The moth larva cannot form the cocoon until the silk glands are matured; that is, this spinning activity is also a deferred activity that cannot be expressed until the moth larva has developed functioning silk glands. Instincts and development are inextricably bound together.

Many writers speak of instinctive and intelligent actions as though these two types of behavior differ fundamentally from one another. Against this view it may be asserted that a large instinctive element enters into every form of intellectual activity; whereas instinctive actions do not usually run their courses altogether automatically and mechanically but contain in addition to fixed and unchanging components a variable element, more or less adapted to the particular situation.

An act, A, is at one and the same time normally a function of the constant, C, and a variable, V; expressed as a formula this becomes

$$A = \int (C \leftrightarrow V).$$

The constant is the instinctive biologically inherited element. The variable is the element which produces, in some cases, an appropriate modified reaction, in others, an unforeseen response to a situation. Analysis of the action into these two components is purely an abstract analysis. V and C must not be taken as two more or less opposed natural agencies pulling the organism now in this direction, now in that, as they battle for supremacy. They are merely two different aspects of the same reality. In inherited, innate, instinctive reactions C is greater than B; in intelligent action the relationship is reversed. I first read this analysis in a book by Alverdes; Yerkes gave a part of the same idea in his 1943 book on chimpanzees.

Two instances of instinctive behavior that I personally observed may be interesting, and, in addition, helpful for the discussion.

Years ago, near the Marine Biological Laboratory at Woods Hole, Massachusetts, I was trotting back from a noontime swim in a hurry for dinner when I saw a solitary wasp dragging an immobilized caterpillar across the uneven surface of the dusty road. I stopped to watch. The wasp turned before she reached the sidewalk. She passed near a low sturdy weed up which she climbed and balanced the caterpillar over a low fork some few inches from the ground. All this time the wasp and her prey had been followed by a small tachina fly, which would lay her own egg or eggs on the egg of the wasp larva. The tachina larva on hatching would feed on the larva of the wasp, while it, in turn was feeding on the paralyzed, living caterpillar.

The wasp descended from the weed, proceeded a few feet to a small burrow, which was a fraction of an inch wide and about that deep. She entered, and enlarged the burrow, digging vigorously.

After a short time the wasp left the enlarged burrow, returned along the way

the most complex of all phases of unlearned behavior, including many levels of behavior from variable appetitive responses to more rigid consummatory actions.

The great confusion about instincts results in part from the fact that much so-called instinctive action of man and other vertebrates, and, perhaps, of some invertebrates also is really behavior based on very early training. The training has been forgotten; the definite activity pattern remains. Sometimes the learning may be based on a single exposure to a given situation.

There are a great many definitions of instinct. These fall into two main types: (*a*) those that are primarily objective, the biologically behavioristic definitions; (*b*) the subjective ones.

Existence of the subjective emphasis is a part of the cause of justifiable skepticism regarding instincts among biologists, including, of course, psychologists. Parenthetically, the scientific part of psychology obviously is a branch of biology. The subjective definitions may be illustrated by the statement that if birds migrate south early in the fall, earlier than they are normally expected to do, this may be taken as a sign that an uncommonly cold winter is coming. The migrating birds are supposed to have a supernatural instinctive insight in the matter. Such fantasies are similar to folktales about whiteness of chicken bones foretelling the severity of the coming winter.

Good definitions of instinct are hard come by. Most of them, even of the objective, biologically behavioristic sort, are involved and difficult to interpret. There are two I wish to present. One is adapted from W. M. Wheeler, who said, approximately, that an instinct is a relatively complicated activity of an organism that is acting (*a*) as a whole rather than as a part; (*b*) as a representative of the species rather than as an individual; (*c*) without previous experience or without modification caused by experience; and (*d*) with an end or purpose of which the reacting animal has no knowledge. More recently, in 1951, N. Tinbergen, in his book entitled *The Study of Instinct*, gives his definition only after he is more than halfway through the book. Even then, the definition is offered almost apologetically in loose association with his admittedly somewhat imprecise discussion of neurophysiological relationships. His statement is (p. 112): "I will tentatively define an instinct as a hierarchically organized nervous mechanism which is susceptible to certain priming, releasing, and directing impulses, of internal as well as of external origin, and which responds to these impulses by coordinated movements that contribute to the maintenance of the individual and the species." These two objective definitions supplement each other more than they overlap. They may be tested against such activity patterns as breathing, swallowing, gland secretion, muscle contraction, all of which may be shown to occur without learning. Some types of glandular secretion, for example, can be effected by learning only with great difficulty, if at all. On casual inspection, such activities as breathing, etc. come more nearly to fitting Tinbergen's definition than they do that advanced by Wheeler, although they run aground there only on the first idea that the organism should be acting as a whole rather than as a part when displaying instinctive behavior.

Instinctive reactions represent examples of prolongation of development exhibited as behavior rather than as formation of new structure. Contrast, for example, the difference between a chrysalid and a cocoon. Each is found at pupation time in the life history of certain members of the Lepidoptera. The butterflies form a chrysalid which con-

sists essentially of a thickening of the hypodermis to form a protective covering. This thickened chrysalid is frequently attached to the walls of some crevice by a gluelike secretion. Inside its protecting covering the pupation processes take place. In contrast, the moths, as pupation time approaches, spin a more or less elaborate silk cocoon. Inside the cocoon the moth larva also develops a somewhat thickened hypodermis, which serves as an additional protection. There is no essential difference in these activities being examples of prolongation of development. One is limited primarily to the hardening of the hypodermis; the other is complicated by the presence of silk-spinning activities which represent instinctive behavior. The moth larva cannot form the cocoon until the silk glands are matured; that is, this spinning activity is also a deferred activity that cannot be expressed until the moth larva has developed functioning silk glands. Instincts and development are inextricably bound together.

Many writers speak of instinctive and intelligent actions as though these two types of behavior differ fundamentally from one another. Against this view it may be asserted that a large instinctive element enters into every form of intellectual activity; whereas instinctive actions do not usually run their courses altogether automatically and mechanically but contain in addition to fixed and unchanging components a variable element, more or less adapted to the particular situation.

An act, A, is at one and the same time normally a function of the constant, C, and a variable, V; expressed as a formula this becomes

$$A = \int (C \leftrightarrow V).$$

The constant is the instinctive biologically inherited element. The variable is the element which produces, in some cases, an appropriate modified reaction, in others, an unforeseen response to a situation. Analysis of the action into these two components is purely an abstract analysis. V and C must not be taken as two more or less opposed natural agencies pulling the organism now in this direction, now in that, as they battle for supremacy. They are merely two different aspects of the same reality. In inherited, innate, instinctive reactions C is greater than B; in intelligent action the relationship is reversed. I first read this analysis in a book by Alverdes; Yerkes gave a part of the same idea in his 1943 book on chimpanzees.

Two instances of instinctive behavior that I personally observed may be interesting, and, in addition, helpful for the discussion.

Years ago, near the Marine Biological Laboratory at Woods Hole, Massachusetts, I was trotting back from a noontime swim in a hurry for dinner when I saw a solitary wasp dragging an immobilized caterpillar across the uneven surface of the dusty road. I stopped to watch. The wasp turned before she reached the sidewalk. She passed near a low sturdy weed up which she climbed and balanced the caterpillar over a low fork some few inches from the ground. All this time the wasp and her prey had been followed by a small tachina fly, which would lay her own egg or eggs on the egg of the wasp larva. The tachina larva on hatching would feed on the larva of the wasp, while it, in turn was feeding on the paralyzed, living caterpillar.

The wasp descended from the weed, proceeded a few feet to a small burrow, which was a fraction of an inch wide and about that deep. She entered, and enlarged the burrow, digging vigorously.

After a short time the wasp left the enlarged burrow, returned along the way

she had come, but did not remember the weed-climbing incident, and did not locate the caterpillar immediately. Meantime the tachina fly had stayed near the caterpillar in place of following the wasp on her digging activities. Finally the wasp located the weed, climbed it, brought down the caterpillar and renewed her slow progress to the burrow. Now the tachina fly followed, keeping, again, about a foot to the rear.

The wasp dug a bit more, dragged the caterpillar into the burrow, and deposited her egg. Just then the tachina fly darted forward, dove into the burrow, presumably laid her own egg or eggs, emerged, and flew away. The wasp filled in the hole, leveled off the ground, and she, too, flew away. All these complicated patterns were examples of innate, unlearned behavior, both by wasp and tachina fly. There was evidence of adjustment to existing irregularities in the surface of the dust, to the size of the burrow, and, also, to the whole weed incident. However, *C* loomed much larger than *V*.

Years later I saw a more elaborate instance of essentially the same behavior —this time out in the mountains in Utah north of Great Salt Lake. There were three of us, including a poet and a psychologist, coming down the mountain trail. We disturbed a wasp carrying her numbed caterpillar enough so that she dropped her prey. We stopped, and she soon actively quartered the region, a few feet from the caterpillar. When she was beginning to waver in her quartering, I picked up the limp caterpillar and deposited it near the wasp, which soon was carrying it on to the partly excavated burrow a short distance away. She repeated the digging and burying routine. The poet sprawling with his eye close to the opening of the burrow saw the glistening egg deposited on the paralyzed caterpillar. Then the wasp filled and leveled off the burrow so that to us the spot was indistinguishable from the bare ground elsewhere except for its being darker from the moisture, which would soon dry. Still the wasp hovered near giving us the impression that she had not yet finished the operation. After a short while she settled down on a fallen cluster of three pine needles still wrapped together as they grew on the tree, transported these to the site of the burrow, and deposited them there. Then that spot looked to us as casually littered as the ground nearby, and the wasp flew away as though entirely finished.

Lorenz, Tinbergen, Baerends, and the other present-day European students of instinct study such behavior patterns by fragmenting them into smaller and smaller units, striving towards breaking them down to their reflex-arc components. At each step they attempt to discover the releasing stimuli, which they call simply releasers. In this they use passive and moving models. They report considerable and promising progress, for example, with the mating activities of grayling butterflies, stickleback fish, and various birds. The analysis of several sorts of instinctive behavior of fishes and birds has been carried fairly far in the identification of sign stimuli that serve as releasing mechanisms, finding, for example, that a bunch of red feathers is the effective releaser of territorial defense for an English robin.

Tinbergen (1951, p. 37) states that the reason why dependence of innate behavior on sign stimuli has not yet been generally recognized notably lies in the fact that so many laboratory psychologists have been studying conditioned reactions. He thinks that conditioned reactions are not usually dependent on a limited set of sign stimuli, but seem to depend on more complicated stimulus situations. The progress these men are making in an obscure phase of behavior is stimulating. It is particularly heart-

ening to see the long stagnant field of instinctive behavior, freighted with "oh my" stories such as I attempted to use in telling about the solitary wasps, beginning to yield to methods of objective analysis.

INSTINCT AS SEEN BY A PSYCHOLOGIST

Henry W. Nissen

As Dr. Allee has shown, many animals exhibit complex patterns of behavior, constant for the species and apparently unlearned, which need a name to set them off descriptively from other behaviors. In the first part of my discussion I shall try to relate the term "instinct," as designation for these uniformities, to certain other concepts, such as reflex, habit, and drive. These considerations will lead to the conclusion that "instinct" has real but limited usefulness, and to the suggestion that all behaviors may be ordered on a continuum of possible associations which are more and less readily learned.

My second point will be a criticism of the current tendency to oversimplify the problem of motivation by organizing all behavior under half a dozen biogenic drives or instincts plus a few psychogenic or secondary drives. This part of the discussion will be illustrated by recent observations of the development of sex behavior in chimpanzees.

Terms referring to the dynamic or energizing aspect of behavior must be kept sharply separate from those pertaining to the form or pattern of the behavior itself. Confusion between the two has been responsible for much of the misunderstanding and controversy about instinct and persists even today: Tinbergen, 1951, presents a diagram of "the hierarchical organization of instincts" which looks like the flow chart of a military chain of command. The top brass—that is, the top instinct—decides the strategy, lower centers direct the tactics, and the lowest-level instincts do the work. The "instinct of pugnacity" refers both to the motivation to fight, and to the manner or pattern of fighting. The term is useful only in the latter sense, which is the way I shall use it here.

The dynamic factor in behavior has been given various names, the most common one, perhaps, being "drive." Morgan has termed it "the central motive state," or CMS, and I have called it the "sensitizing component of behavior determination." The drive or sensitizing factor elicits behavior which is often such as to increase or decrease the amount of a certain kind of stimulation. Thus the earthworm may orient away from the source of light, whereas the Euglena may swim towards the light. Instead of deriving from the external environment, the stimulation may arise internally: the organism sneezes, defecates, or urinates. In other cases, the effective condition is a surplus or deficiency of chemicals in the bloodstream, presumably sensitizing certain parts of the nervous system. In still other cases, the mechanism of the energizing factor is completely unknown. You will note that I am postulating a sensitizing factor in all behavior, whether it is of external or internal origin, and whether it leads to behavior as simple as a reflex or as complicated as nest-building.

Now the behavior which ensues when a sensitizing factor is active may differ in various ways. Most obviously, it may be all over in a moment, or it may last a long time. When the goal or situation which brings relief is immediately present, the response is prompt and brief; we call it an unconditioned or conditioned reflex, tropism or taxis, or automatized habit. When the goal object or situation is not at hand, a more or less prolonged series of acts occurs. Duration in time is therefore one differentiating criterion. Further, a reflex typically is elicited by a specific, local-

ized stimulus, whereas an instinct is usually determined by a combination of internal factors plus a *class* of external stimulus patterns which, in their details, may vary considerably. On the response side, a reflex implies contraction of a particular muscle or muscle group, whereas instinctive behavior allows considerable leeway in the effector mechanisms by which a common effect is achieved.

In the absence of the goal or drive-reducing situation, the organism is forced to an indirect, extended series of acts. These longer sequences may be classified into five categories, in accordance with information available to us regarding *how* the form of that behavior was determined: (*a*) When little or no direction or selectivity is in evidence, we speak of "random" or "spontaneous" activity. (*b*) When innately determined components are conspicuous, we call the sequence an "instinct." (*c*) When learning has determined the pattern, we call it a "complex habit." (*d*) When there is evidence of reasoning and foresight, we speak of "purposive striving" or "goal-directed" activity. (*e*) When we want to remain neutral about the role of the innate, experiential, and central factors involved, we call it simply "drive behavior."

Together with the other terms designating prolonged behavior sequences, "instinct" implies the sensitizing effect of neural stimulation. An animal behaves differently when hungry than when thirsty. The drive or motivating factor sensitizes the animal to some stimuli and makes it obtuse to others. It both filters and intensifies. This selective action does *not* differentiate among the five classes of long sequences. Nor does the consummatory response, which usually gives the behavior its name, and which may be the same for random, instinctive, rational, and habitual responses. It is in the etiology of the behavior preceding the consummatory response, in the past history of the animal, that the differentiating criteria are to be sought. This demands experimental analysis. Sometimes the data are unambiguous, as when the isolated bird sings the song which is characteristic of its species. Both birds and chimpanzees build tree nests; in the former the behavior pattern is often innate, but in the apes it is evidently learned, being transmitted from one generation to the next. Often the answer is complicated, because the behavior sequence contains both innately determined and learned components. Instinct and intelligence are *not* mutually exclusive. Some of the fastest and most efficient learning that we know of is intimately related to instinctive behavior: the foraging insect must learn and remember, on the basis of a single flight, the direction or the landmarks which guide it back home. According to Baerends, the digger wasp even remembers from day to day how well each of its eight or ten burrows is stocked with provisions.

Finally, I should like to suggest that reflexes, instincts, and the inherited capacity to learn may be distributed on a continuum. What is inherited may be *a more or less specific readiness to learn*. The concept of learning implies that such readiness is rather nonspecific. But often, as in the case of the wasp, there is a readiness to learn very specific things. When such selective readiness is common to the species, the resulting behavior can hardly be distinguished from instinct or reflex.

We come, now, to a criticism of the tendency to ascribe all or most behavior to a few drives or instincts. During the past two years I have been observing the behavior of male-female pairings of late-adolescent chimpanzees. As youngsters these animals lived together, but well before puberty they were separated by sexes. Starting at least a year after the first menstruation, and at a higher

chronological age for the males, these animals were put together for periods of observation at times when the female was in swelling, that is, during the periods of physiological receptivity. All possible pairings of five sexually naive males and a like number of females were made. The observations total over 100 in number. If I had used experienced, rather than inexperienced, chimpanzees under like conditions, a conservative estimate of the number of copulations which would have occurred is 200. In my observations there were no copulations.

However, except for what Carpenter has delicately called "primary sex activity," the chimpanzees were very active during these hours of observation. These primates are notoriously inventive and varied in their behavior, and practically everything that a caged ape *can* do these animals *did* do in the course of my observations. The behaviors which occurred were both individual and social. The former included self-grooming, solo gymnastics such as somersaulting and doing cartwheels, sucking a thumb, or just sitting and thinking. Among the social interactions there were a few instances of serious aggression, many occurrences of bluffing or exhibitionistic behavior, a great deal of playfighting, wrestling, playful slapping or boxing, tag or follow-the-leader, and mutual grooming. The point to be stressed is that, although there was no copulation, there was a great deal of social behavior, including most of the constituent acts which enter into the mating pattern.

The question now arises whether the observed behavior which did occur is properly and appropriately designated as sex behavior, as expression of the sex drive or instinct. The only excuse for doing so is *our* knowledge (*a*) that with other, experienced, male-female pairings the sex act would occur, and (*b*) that, with continuation of the prevailing conditions for a long enough time, the sex act most probably *will* occur, eventually, in each of these 25 pairings. But this is not adequate justification for ascribing these long sequences of variable behavior to the "sex drive." There was no consummatory act and no drive reduction; the reactions were not "preparatory" (except in a farfetched teleological sense) for mating.

What, then, did motivate the great variety of behavior which was observed? There are several theoretical possibilities. The old concept of youthful play as being the incomplete, nonserious expression of later biologically significant behavior patterns is particularly unsatisfactory here because, structurally and physiologically, these animals *are* ready for actual mating and its consequences. A second possibility is to postulate an independent drive, such as activity, exploration, play, or curiosity. The conceptualizations are too glib and facile—they "explain" too much too easily, and give no "handle" for experimental testing. Still another possibility is the idea of displacement reactions, as proposed by the German neonaturalists. They suggest that when the environment does not provide the stimuli or objects necessary for the development of the currently dominant instinct, reactions belonging to some other instinct will occur. This also is an *ad hoc* explanation which can be brought in as necessary to "explain" almost anything.

Instead of starting with a dozen or so drives, instincts, or propensities, under one or another of which all behavior is ordered, we may, instead, postulate a multiplicity of self-motivated activities. What were conceived of as part-activities, all contributing towards some one definite end, are thought of, instead, as a series of independently motivated acts. Every postural adjustment, every approach to or avoidance of a given ob-

ject, each episode of grooming, has its own, intrinsic motivation. To explain the vast number of movements and acts, extending over hours, and sometimes over weeks and months, as all being determined and guided by one drive, whose direct and identifiable expression can be seen only in a brief consummatory act, is pure anthropomorphism. The acrobatics, wrestling, and grooming seen in my male-female pairs of chimpanzees cover a longer period of time, and involve a greater variety of precise coordinations, than does the act of mating which takes less than a minute. But mating is all that may be strictly designated as sex behavior.

Mutual grooming by chimpanzees contains elements which, in human behavior, we call foreplay or petting and which we ascribe to the sex drive. Could it be suggested that petting is, sometimes, an end in itself, with no sinister motive towards a further ulterior goal? In chimpanzees, grooming occurs as frequently in preadolescent animals, who have had no sex experience, as it does in adults; it occurs in female-female pairs as much as in male-female pairs; and it occurs more often after mating than before. Looked at from the outside and as a whole, the goal-directedness of the behavior sequence is obvious, but that is the view seen by the human mind, not the view of the organism doing the behaving. To suppose that it is, is to endow the animal, anthropomorphically, with the foresight which characterizes the deliberate, devious, and planful behavior of man.

The behaviors legitimately and descriptively named sex, hunger, thirst, and so on, are relatively infrequent, isolated events in the flow of behavior; their motivation demands explanation no more, and no less, than do the many activities appearing in each of various sequences. To say that a given act is sometimes motivated by sex, another time by hunger, is to slur over the basic question of motivation. Differential sensitization, determining the probability of occurrence of various reactions, needs explanation. But more fundamental than the problem of frequency of elicitation, is *why* grooming and wrestling and play-biting occur at all. Since they appear when there is no copulatory drive, and more often after reduction of the sex drive than before, they must be independently motivated.

The theme of the view which I am advocating might be summarized in the words of a once-popular song, whose title is, "Every little movement has a meaning all its own." I should perhaps restate this to read, "Every little action has a motive all its own." My point, of course, is not that sex has been overrated, but rather that it already has enough to account for, without our burdening it with more than its fair share of responsibility.

A SOCIOLOGIST'S VIEW OF INSTINCT

Meyer F. Nimkoff

Since Dr. Allee, a zoologist, has dealt mainly with the lower animals, and Dr. Nissen, a psychologist, with chimpanzees, my function is to consider the question of instinct in relation to man. Although sociologists concern themselves primarily with man, they need for proper perspective on this question of instinct some knowledge of the motivations of other animals. This knowledge comes largely from zoology and psychology. Sociologists may, of course, properly study the social life of animals other than man, although few do so.

To avoid semantic difficulties, it is desirable at the outset to clarify the concept of instinct, which the two preceding discussants have done. Central to the concept are the ideas that the behavior is complex, common to the species, and unlearned. An example is the

building of nests by birds. A bird that has never seen any other bird build a nest will still build a nest if the conditions are right. There is, in fact, an inner compulsion that forces certain animals through the motions of nest-building even when the necessary materials are lacking. The nests of different species of birds may vary much more than do the nests of any one species.

Does man have any instincts? The early psychologists, William James, William McDougall, and E. L. Thorndike, drew up long lists of instincts. James said man had more instincts than any other animal. His list included sucking, crying, locomotion, curiosity, shyness, cleanliness, pugnacity, fear of dark places, acquisitiveness, love, jealousy.

About 1924 the reaction set in with the publication of *Instinct* by L. L. Bernard, a sociologist, who showed how confused was the use of the term. Covering some 400 authors, he disclosed that about six thousand urges had been called instinctive. These were of two kinds, those of a general nature like sex and social behavior, and those that were more specific like "an instinct for the piano" and "an instinct to avoid eating the apples of one's own orchard." From our perspective today, Bernard's analysis seems like a caricature. Instinct in human psychology was buried, and Knight Dunlap and John B. Watson added their nails to the coffin with their accent on conditioning and learning.

The tendency in recent years has been to reserve the term instinct to the lower animals. Man's nervous system is not as imperious in dictating behavior because it is more complex than that of other animals and is less fully coordinated at birth. For instance, it takes a human being longer to learn to walk than it does an ape. This complexity and plasticity of the human organism makes possible more and richer learning. So, in the case of human beings, we drop the term instincts and speak of organic drives, which are more general. There are also drives of a purely social origin which have been termed wishes or motives. Most of man's behavior is learned, whereas most of a bird's behavior is not learned.

The problem of instinct appears to be a matter of differences in the degree of the learning capacity of organisms. Some students object to the use of the term instinct even in reference to the lower animals because they, too, may be capable of learning and are not mechanistically repetitive in their behavior. Even the lowly amoeba can, apparently, learn to alter its behavior in response to new stimuli. Birds improve the quality of their nests with practice, and learn from other birds. So the differences between man and the lower animals can be exaggerated. The anthropomorphic danger is ever present, and man must beware of his bias in favor of himself. The evolutionary viewpoint is a wholesome corrective, emphasizing as it does differences in degree as well as differences in kind.

We can stress the fact that differences in learning ability fall along a continuum, but this need not cause us to lose sight of the fact that differences in degree may be very considerable indeed. The ants have complex behavior patterns based on the division of labor, differentiation of status, and different classes of workers. These patterns are inherent in the structure of each ant and show little change except as the ant may change. Professor W. M. Wheeler examined ants embedded in Baltic amber fifty to seventy-five million years ago and concluded that these ants had developed all their various castes just as they exist today. The larvae and the pupae were the same. They kept guest beetles in their nests and had parasitic mites attached to their legs in the same special positions as do our

species today. Apparently the ants have learned very little in fifty million years. On the other hand, man has learned much. A community like New York City with its eight million inhabitants and complex social organization is very different indeed from an Eskimo band of 10 to 100 individuals or thereabouts, a change which represents a cultural span of about 15,000 years. So we may say that while other animals may have some culture, the amount is negligible, and man is the only animal with a substantial and significant culture. The differences in cultural level among animals reflect differences in learning capacity, a biological phenomenon. Among large groups of men, like races and nations, differences in learning are the result of differences in opportunities for building culture, not differences in inherited learning ability.

Sociology depends on biology and psychology for knowledge of the inherited biological nature of man. What can sociology itself contribute to our understanding of this problem? Sociology is interested in the structure and function of social systems: that is, the total cultural organization that constitutes man's social environment. We cannot control the human environment experimentally, like that of birds or chimpanzees, to observe how individuals behave when reared in isolation. The folkways and mores, which are so important to human learning, nowhere permit this.

While it does not seem possible to hold the human environment constant by isolating individuals and denying them the opportunity of learning from the group, we can in a sense reverse the situation and observe what happens when the genetic factor is held constant, as it is when very large groups of human beings are involved. Individual differences in genetic traits are minimized or cancelled out under these conditions.

What we find in different human societies is a considerable diversity of customs. Some cultures are strictly monogamous, others polyandrous, still others polygamous. Some societies do not permit divorce, most societies frown on it but permit it, and a few encourage it. The folkways are marked by variety. This is thought to demonstrate man's highly flexible inherited tendencies.

Underlying the diversities, however, are certain uniformities. Whether it be monogamous, polyandrous, or polygamous, the family is everywhere to be found. So it is natural to ask: do the uniformities in culture the world over reflect certain common human drives that are genetically determined? Are the cultural universals an expression of man's unique biological nature?

In the case of the family, the answer seems clearly to be in the affirmative. The universality of the human family as a social institution rests on the fact that there are two sexes endowed with a sex drive and that the human infant is highly dependent on others for survival. The family is omnipresent, but the organization of the family varies greatly. So we conclude that the presence and broad outlines of this institution are determined by man's biology, but not the details of structure and function.

Cultural universals are not the only indicators of man's innate biological nature. Cultural preferences may be indicators, too. While culture patterns vary widely, some patterns are more common than others. Polygamy, the marriage of one male and several females, is much more frequent than polyandry, the mating of several males and one female. Does this suggest that polygamy is more compatible with the biological nature of man than polyandry? The evidence from the primates supports this interpretation. The male, generally larger than the female, often shows signs of jealousy and possessive-

ness. The organization of a baboon family is described by Zuckerman as consisting of a male overlord, his female or females, together with their young, and sometimes including one or more "bachelors" or unmated males.

It is not safe, however, to infer that behavior is genetically determined simply because it is found to be widespread, or even universal. Take, for example, the taboo against incest. Every society enjoins incest, although the definition varies, as does that of the family. Violation of the incest taboo is severely punished, and few if any offenses are regarded as more serious. Is the horror of incest "instinctive," as some have argued? If so, it would hardly seem necessary to have a taboo against it. The explanation of the taboo is not clear, as may be gathered by an examination of the voluminous literature on the subject. A more plausible explanation for the incest taboo is that it is the result of universal social factors and that it helps to promote social organization. It is difficult, for instance, to see how authority patterns could be maintained in the family without it. Therefore we are unable to conclude that human behavior which is universal must, by virtue of this fact, be biologically motivated. There may be universal considerations in the acquired social situation which prompt the behavior.

Moreover, behavior which is biologically induced may be modified by the culture so that its usual or normal expressions are not evident. The sex drive, for instance, can be repressed. People can be persuaded to fast and even starve to death, although it is difficult to see how the latter behavior could be universalized if society is to persist.

In short, there does not seem to be any easy formula which we can use for determining what is learned and what is genetically motivated in human behavior.

[MS. received July 7, 1952]

Psychological Review
Vol. 60, No. 5, 1953

ON THE PROBLEM OF PERCEPTUAL DEFENSE

LEO POSTMAN

University of California

In a recent issue of this journal, Howie (9) published a spirited attack on the concept of perceptual defense used by Bruner and Postman in the analysis of motivational factors in perception (2, 3, 4, 20). Since the writer had himself come to the conclusion that the facts which gave rise to this concept can be better subsumed under other, more general, principles of perception (16, 18, 21), he would have been ready to welcome Howie's statement. It appears, however, that Howie was right for the wrong reasons and used questionable polemics to make his points. The purpose of the present note is to reconsider the status of perceptual defense and to clear up some of the misconceptions which appear in Howie's argument.

Origins and Status of the Concept

Original statement of the concept. Let us begin by briefly sketching the development and present status of the perceptual defense concept. The principle was first formulated in connection with the analysis of recognition thresholds for words from different meaning classes. It was found that negatively valued stimuli had, under some conditions, higher thresholds than positively valued or neutral ones. In addition, the prerecognition responses to the negative stimuli were such as to suggest avoidance of their recognition (20). These findings led to the speculation that the recognition of negatively valued stimuli is delayed by "perceptual defense mechanisms." It was acknowledged and, indeed, emphasized from the very beginning that *perceptual defense was not in itself an explanatory principle* and that the mechanisms mediating it remained to be discovered. To quote from the first statement concerning perceptual defense: "One may inquire at this point, 'How does the subject know that a word should be avoided?' In order to repress, he must first recognize it for what it is. We have no answer to propose. . . . Of only one thing we can be fairly sure: reactions do occur without conscious awareness of what one is reacting to. Psychological defense in perception is but one instance of such 'unconscious' reaction" (20, p. 152). Clearly, then, perceptual defense was a descriptive principle of perceptual functioning—a statement about an outcome of threshold measurement—whose specific mediating mechanisms were a problem for the future.

The problem of mediating mechanisms. The next step was to speculate about possible mechanisms (4). Hypotheses about mediating mechanisms had, of course, to come to grips with the apparent paradox of perceptual defense—that the stimulus has to be somehow discriminated in order for the processes delaying its recognition to be brought into play. Upon closer inspection the paradox turned out to be a semantic problem. There may be more than one kind of discrimination, in the sense that more than one class of responses may be systematically related to variations in the stimulus. Such different discriminations may (*a*) have different thresholds and (*b*) facilitate *or* inhibit each other. It has been shown, for example, that reduced stimulation insufficient for correct verbal report (such as a brief tachistoscopic exposure) may be adequate to trip off conditioned

autonomic responses (12, 15).[1] Such autonomic responses (discrimination without awareness) may in turn delay the appearance of the correct verbal response defining discrimination with awareness. Exactly this sequence of events was proposed by McGinnies (15) in explanation of what appeared to be defensively high thresholds for the recognition of taboo words. I am not here arguing in favor of that specific hypothesis; in fact, as I shall try to show presently, other conceptualizations of these data are more satisfactory. What I wish to insist on at this point is (*a*) that perceptual defense referred to an observed property of certain recognition thresholds, (*b*) that there was nothing inherently illogical about such a concept, and (*c*) that it was possible to hypothesize specific mechanisms mediating it.

Reanalysis and reduction to general principles. The concept of perceptual defense did, however, prove an uneconomical one, or at least an expendable one. What had appeared as a special principle could, upon reanalysis, be considered a special case of more general principles of perceptual functioning. There were two sets of considerations which led to this conclusion: problems of experimental control, and theoretical parsimony.

1. Experimental control. Empirically, perceptual defense refers to the *difference* between recognition thresholds for neutral and anxiety-arousing stimuli. Clearly, perceptual defense can be inferred on the basis of such data only if other more general determinants of the observed differences in thresholds have been adequately controlled. Such control has proved difficult. For example, the striking differences between recognition thresholds for neutral and taboo words reported by McGinnies (15) can upon reanalysis be ascribed largely to differences in familiarity between the two classes of words and the effects of selective-reporting sets (7, 18). In other studies in which complex stimulus materials such as sentences (23) and pictures (5) were used, the evaluation of the relative familiarity and structural difficulty of the neutral and critical stimuli is so difficult that the threshold differences cannot be interpreted as evidence for a defensive process. Similar difficulties of interpretation arise in studies which correlate sensitivity to critical stimuli with independent personality measures of the *S*s (5, 6, 11). Even though the personality measures purport to classify *S*s in terms of their proneness to defensive reactions, it does not follow that the threshold differences observed do, in fact, represent different degrees of perceptual defense. A rigorous operational definition of *defense* would have to anchor the concept in *antecedent* as well as in consequent conditions. Since the antecedent conditions have not been fully specified or brought under experimental control, alternative interpretations of the observed threshold differences in terms of such variables as familiarity and selective set remain possible. Such alternative interpretations are indirectly supported by the fact that variables like familiarity and selective set *can* be experimentally manipulated and do produce variations in thresholds analogous to those ascribed to perceptual defense (18).

2. Theoretical parsimony. In the development of a general theory of perception one of the major requirements is the formulation of constructs which have the highest possible degree of generality. The applicability of such concepts should, for example, not be restricted to

[1] Before such autonomic responses can be confidently regarded as evidence for systematic discrimination rather than as artifacts of the method of measurement (cf. 7), they should be validated by means of other classes of responses (cf. 17).

special types of motivational situations. Tentative attempts at the formulation of such a theory, centering about the concept of *perceptual hypothesis,* have been published elsewhere (1, 16). A hypothesis was defined as a predisposition of the perceiver to organize stimulus cues in specific ways. Such hypotheses are anchored on the antecedent side in conditions of stimulus input and specified conditions of the organism (including drives and motives) and on the consequent side in systematic perceptual responses (discriminations, verbal reports, etc.). Hypotheses vary in strength, i.e., they vary in the amount of stimulation necessary to arouse, confirm, or deny them. We have found the concept of hypothesis useful in the analysis of a considerable body of experimental data, including the phenomena of perceptual defense which can now be economically conceptualized in terms of interference among competing hypotheses. "What appears to be perceptual defense results from the dominance of strong alternative hypotheses rather than from active repression of the inimical or dangerous. In the presence of partial information, strong hypotheses incompatible with the threatening stimulus may be evoked. . . . If this is the case, the subject will appear to be defending himself against perception. . . . If, however, hypotheses related to the negative stimuli are strong, the opposite of defense will appear to operate" (16, pp. 256 f.).[2] Thus, the demands of theoretical parsimony led us to abandon perceptual defense as a special principle of perception and to fit the phenomena to which it referred into a broader theoretical context.

Such, as we see it, is the "case history" of the concept of perceptual defense to date. The concept was suggested by challenging experimental data and served the purpose of raising useful empirical questions and sharpening theoretical issues. *The reasons for which such a concept is accepted or rejected* appear to us to be of critical importance, for they reflect the intellectual climate in which theory construction proceeds, the rules by which we agree to abide in our theorizing even though specific constructs may come and go. I believe that Howie's critique of perceptual defense is largely based on criteria whose acceptance would be dangerous to sound psychological theorizing and to experimental initiative. Thus, even though I do not now hold a special brief for the concept of perceptual defense, I feel it necessary to register my disagreement with most of his arguments.

EVALUATION OF HOWIE'S CRITIQUE

Several lines of criticism are advanced by Howie in his attack on perceptual defense—operational, logical, and, most important and pervasive of all, metaphysical. Let us now consider them in detail.

Relation between defense and other motivational principles. The principle of perceptual *defense* asserted that the perceptual process serves to minimize recognition of objects and events inimical to need and expectations. This principle was supplemented by the apparently antithetical principles of perceptual *vigilance* and perceptual *resonance* (2, 20). Vigilance referred to the empirical observation that under some conditions the organism is more sensitive to threatening or inimical stimuli than to neutral ones. Resonance referred to *S*'s tendency to give prerecognition hypotheses (i.e., to interpret reduced, un-

[2] This shift in interpretation from an inhibitory mechanism to a principle of competition among incompatible responses parallels developments in the theory of retroactive inhibition (cf. 14). It is to be noted that in his paper Howie makes no reference to the discussion quoted here or to a related statement (21) bearing on the interpretation of perceptual defense.

clear stimuli) in accordance with his needs and expectancies. As a result the probability of an early correct recognition response is increased if the stimulus actually belongs to the class of objects (e.g., words) which are the major sources of prerecognition hypotheses. The interrelation among these three principles provides Howie with his first target of attack. The criticism centers around two points: (*a*) the difficulty of operational delimitation of these antithetical concepts and (*b*) the imputed motivation behind their postulation.

Let us turn to the operational criticism first. The legitimate question is raised as to where one principle begins and the other ends, i.e., how these different concepts are to be concretely applied to empirical data (cf. also 13). It is true that some of the early statements of the three principles made their precise demarcation difficult when they were used in the speculative interpretation of observed perceptual responses. They can, however, be given adequate and mutually exclusive definitions. Let us remember that the concepts were introduced to account for observed differences in thresholds for stimuli from different meaning classes. Using the thresholds for standard "neutral" stimuli as points of reference, thresholds for stimuli from specific meaning classes which are significantly higher would define defense, thresholds which are significantly lower would define vigilance. These specific meaning classes would, of course, have to be defined *independently of the threshold data* on the basis of a theoretical analysis of the conditions of defense and/or vigilance. Ideally, the critical meaning characteristics should be produced by means of experimental manipulations of the organism, such as the establishment of a conditioned fear response (cf. 12). The critical classes of stimuli would also have to be equated to the standard stimuli in terms of other characteristics (familiarity, structural properties) relevant to their discriminability. Short of such control of antecedent conditions, alternative explanations of the threshold differences are possible and a precise demarcation of significantly "vigilant" and significantly "defensive" thresholds must remain difficult.

The situation is, of course, *in no way different* for the concepts of facilitation and inhibition in learning which Howie contrasts favorably with the defense-vigilance antinomy. When the learning of a list A has a significant influence on the learning or retention of list B, it is often impossible to tell to what extent facilitation and inhibition both entered into the transfer effect. Facilitation and inhibition must be defined in terms of the net transfer effects obtained under standard experimental conditions.

As for resonance, it is capable of direct identification and measurement in terms of the latency and frequency of different types of prerecognition hypotheses given by *S* prior to correct recognition of the stimulus. The less the latency and/or the higher the frequency of prerecognition hypotheses from the same meaning class as the stimulus, the higher is the degree of resonance. The same feature of perceptual responses has been analyzed by Solomon and Howes (8, 24) under the heading of response probability.

In summary, then, descriptive definitions allowing for the demarcation of defense, vigilance, and resonance are possible. As was pointed out above, however, theoretical parsimony leads to the abandonment of these as principles in their own right since they are reducible to special cases of more general principles.

Howie directs his major criticism at what he believes to be the reasons for the introduction of antithetical principles such as defense and vigilance. In

discussing the principle of vigilance he writes, "In this notion we have an attempt to incorporate within the theory of perception a reality principle without which resonance and defense would make veridical perception impossible and, confining the person within his economy of need determinations, would deny him any effective adjustment to conditions other than his wishes or his fears" (9, p. 309). And in another place he avers that vigilance is "brought in to meet the ultimate solipsistic difficulty of resonance and defense-determined perception" (9, p. 310). These reasons are Howie's; they were not ours. The introduction of these principles stemmed from empirical observations, not from a struggle with problems of epistemology and metaphysics. Perceptual phenomena are what they are, no matter what may be the nature of the "real," concern with which we are content to leave to the philosophers. However, if we may allow ourselves for the moment to engage in epistemological argument, it can be easily shown that a principle of perceptual defense does not involve us in the "ultimate solipsistic difficulty" from which we must then extricate ourselves by means of a principle of vigilance. Perceptual defense refers to observed *quantitative variations* in recognition thresholds, i.e., it was asserted that within the constraints of the stimulus situation the individual will avoid or delay recognition of inimical stimuli. It has never been asserted that the individual is incapable of recognizing inimical stimuli and hence is in danger of being encapsulated in an autistic world of his own; the only claim was that the thresholds for some classes of objects under some conditions of the organism were higher than for other classes of objects. By the same token, the principle of vigilance denoted the fact that under some conditions of the organism thresholds for some classes of objects are lowered.[3] As a statement concerning variations in sensitivity, perceptual defense involves neither greater nor lesser epistemological hazards than do such time-honored psychological stand-bys as "selective set" and "attention."

It is interesting to note that Howie is perfectly willing, as are we, to speak of the determination of perceptual recognition by "conditions of relative set-dominance." But then, if we are to fall in with his epistemological worries, is there not serious danger that, to paraphrase the statement previously quoted, the individual will be "confined within the economy of his set-dominances which would deny him any effective adjustment to conditions other than those fitting his set-dominances"? Selective sensitivity is still selective sensitivity by any other phrase and is, moreover, an empirical fact, whether or not it fits any epistemological preconceptions. It appears that the objection cannot be to the hypothesis of perceptual selectivity but rather to the word, "defense." And an ill-chosen word it was if unwarranted surplus meanings and overinterpretations of the term are as compelling as they seem to be.

The logical paradox of perceptual defense. Howie's second major criticism concerns the logical paradox implied by the concept of perceptual defense: "To speak of perceptual defense . . . is to speak of the perceptual process as somehow being both a process of knowing and a process of avoiding knowing" (9, p. 310). We have already considered this problem and have tried to show that it is a semantic one arising out of ambiguities in the definition of the term *knowing*. A few additional comments are in order here. It seems quite inap-

[3] The statement that certain classes of discriminations will be "minimized" or "maximized" (4), to which Howie objects, is to be interpreted in this sense only.

propriate to bar the use of a concept by constructing verbal contradictions which somehow seem to be suggested by the concept. The only contradictions which provide serious grounds for objection are incompatible operations or incompatible deductions entailed by the use of a given concept. The context of empirical investigation in which perceptual defense was used made it clear from the very beginning that it referred to a lack of one-to-one correspondence between two different sets of perceptual discriminations—those measured by correct verbal response and those inferred from other systematic response tendencies in the presence of the stimulus. What is gained by then labeling both sets of discrimination as *knowing*, and announcing that the concept entails a simultaneous process of knowing and avoiding to know? [4]

Finally, it should be noted that Howie takes cognizance, in passing, of the resolution of the paradox (9, p. 310, footnote 2) but then dismisses it and proceeds with his criticism since this resolution "involves a shift in status for perceptual defense. Instead of a basic principle of perception, it has become something to be explained away." The shift in status is of Howie's own construction. As we have pointed out repeatedly, perceptual defense was not and could not be considered an ultimate irreducible principle which did not require further analysis in terms of mediating mechanisms.

The motivation behind motivational concepts in perception. Having pointed to the operational, logical, and epistemological flaws in the concept of perceptual defense, Howie addresses himself to the motivation behind the acceptance of this concept. "Why," he asks, "should it show such vitality, why should it prove so insidiously attractive?" (9, p. 311). To analyze the intellectual motives of one's colleagues is a hazardous venture indeed, but Howie believes that there are two main sources of confusion which account for the insidious seductive power of the concept of perceptual defense: "(1) the interpretation of processes in terms of agency concepts and (2) the internalism of an exclusive emphasis on impulse" (9, p. 311).

I am still not too clear about the first of these sources of confusion since it is stated in only the most general, and nebulous, terms. To speak of perceptual defense is said to "commit a category mistake in representing 'the facts of mental life as if they belonged to one logical type or category . . . when they actually belong to another' " (9, p. 311). As far as I can gather, the main objection is to the ascription of defense to the perceptual process rather than to the person as a whole. Persons may perhaps defend themselves against perceptions, e.g., by shutting their eyes, but perceptual processes cannot. If I read the argument correctly, I submit that we are dealing with a semantic strawman which was erected only to be annihilated. When we ascribe properties to a process and then in turn attempt to explain certain observed outcomes in terms of these properties, we do not thereby endow the process with the status of an agent with aims of its own, functioning independently of the total organism.

[4] The inappropriateness of such play on words may be further illustrated by an example from learning theory. One of the fundamental assumptions of Hull's theory is that the continued reinforcement of a stimulus-response connection produces not only increments in habit strength but is also conducive to conditioned inhibition, i.e., the acquisition of a habit not to respond. Hull carefully distinguishes the conditions and quantitative properties of these two simultaneous processes and the hypothesis has been experimentally fruitful. Yet a superficial semantic argument of the type discussed here would give short shrift to Hull's hypothesis by asserting that reinforcement cannot at the same time be a condition of learning to respond and learning not to respond.

In the interests of analysis, we identify certain processes, such as perception, and try to give as full a description of their rules of functioning as possible. If we were to take Howie's injunction seriously, it would be equally inappropriate to analyze the memory process in terms such as retroactive and proactive inhibition, for after all, it is the whole man, and not the memory process, which does the remembering and forgetting! As long as the definition of processes remains anchored in antecedent and consequent conditions, we need not aggravate our problems with worries about their status as independent or dependent agents.

The second alleged source of confusion is the "tendency to explain behavior on a basis of organic impulse in fundamental separation from external influences" (9, p. 311). In considering this criticism, we must raise two separate and distinct questions: (*a*) is such a tendency necessarily implied by a concept such as perceptual defense (or, in general, by the hypothesis that needs may influence perception) and (*b*) has there, in point of actual fact, been such a tendency on the part of certain investigators? I believe the answer to the first question is no, the answer to the second question is yes.

Operationally and empirically, there is nothing in the concept of perceptual defense which entails an interpretation of behavior on the basis of organic impulses in separation from external influences. As was pointed out above, defense referred to quantitative variation in recognition thresholds. Raised thresholds for certain types of stimuli were designated as defensive. This means that conditions of the organism may modulate or affect the perceptual process and, in this case, cause the final correct discrimination to be delayed. In the same context, we have often spoken of the *joint determination* of perceptual discrimination by sensory and directive factors (4, 16, 22).[5] Such an analysis cannot reasonably be construed as representing an interpretation of behavior exclusively in terms of motivational determinants without regard to external (stimulus) determinants. Indeed, the very operations of the experiments in which perceptual defense was believed to have been demonstrated—the measurement of discrimination thresholds for certain classes of stimuli as a function of such variables as exposure time and/or intensity of illumination—clearly disposes of the allegation of internalism.

It is true, however, that some investigators have, in the interpretation of their results, shown a "projective bias," i.e., they have neglected to carry out a full and precise analysis of the stimulus conditions and response dispositions on which their inferences about perceptual discrimination were based.[6] Such a projective bias easily leads the investigator to ascribe to the operation of special motivational mechanisms observed variations in perception which can be economically interpreted in terms of more general principles of discrimination (cf. 7, 18). There is nothing about the hypothesis that needs may influence perception which, of necessity, leads to a projective bias in the interpretation of experimental data. In each experimental investigation, it is an empirical question whether and to what extent the manipulation of motivational variables results in significant effects upon perceptual response dispositions.

The final argument in Howie's critique attempts to interpret the hypothesis of perceptual defense as an im-

[5] Cf. also the distinction made by Krech and Crutchfield (10) between structural and functional factors.

[6] For a discussion of an operational analysis of experiments on motivation and perception see (17).

plicit endorsement of Freud's principle of economy—that the organism has a basic drive to protect itself against stimulation. The chain of reasoning which leads to this rather startling conclusion is somewhat as follows: (*a*) a "suspicion" that the principle of perceptual defense is an offspring of the economy principle, (*b*) *ergo,* the principle of perceptual defense partakes of the same difficulties as does the principle of economy. By a series of assumptions and extrapolations which leave one rather breathless the hypothesis of perceptual defense is made tantamount to the theory of the death instinct. And a hypothesis which was designed to account for certain variations in recognition thresholds now has the consequence that "adjustment, intelligent behavior, and achievement become but 'more complicated and circuitous routes to the attainment of the goal of death' [*sic!*]" (9, p. 314).

It seems hardly necessary to try to refute this argument in detail. I would venture only the following remarks.

1. Howie's suspicions about the ancestry of perceptual defense were unfounded; he worried needlessly.

2. Although I am as much opposed to some of the excesses of Freudian theory as anyone else, I would hope that it is possible today to subject concepts associated in one way or another with psychoanalysis to experimental tests without arousing irrelevant emotional reactions to the entire body of Freudian doctrine.

3. The evaluation of concepts in terms of the motives which allegedly led to their introduction cannot but result in regrettable misunderstandings and futile debate. Concepts must stand or fall in their own right, whatever the reasons, motives, or hunches which led someone to think of them. And in any event, however astute the critic, he is as likely as not to fail in his attempts at reconstructing the thought processes of the theorists with whom he wishes to disagree.

Summary and Conclusions

The concept of perceptual defense was formulated to account for a specific set of empirical data—raised recognition thresholds for stimuli believed to be anxiety-arousing. The concept was not, and cannot be, regarded as an irreducible explanatory principle. To be maintained, it would have to be anchored in antecedent and consequent conditions, and the mechanisms mediating defense would have to be specified and related to other principles of perceptual functioning. Experimental and theoretical reanalysis suggests that perceptual defense need not, at least for the present, be regarded as a special principle of perception. The data to which it refers can be conceptualized in terms of more general principles.

The concept cannot, however, legitimately be rejected on purely logical, epistemological, or metaphysical grounds.

The logical paradox of "knowing in order not to know" disappears when it is recognized that more than one class of discriminatory responses may occur in response to the same stimulus. Such different discriminations need not have the same thresholds and may under some conditions inhibit each other.

It is inappropriate to evaluate hypotheses concerning perceptual selectivity in terms of preconceptions about a required degree of correspondence between the "real" and the perceptual. How well perceptual discrimination corresponds, in any given situation, with other measurements of the stimulus is an empirical question.

The hypothesis that motivational factors may influence perception (e.g., result in delay of recognition) does not entail neglect of stimulus determinants or a disregard for the *S*'s verbal and

motor response dispositions. Such a hypothesis merely focuses attention on the contribution of additional variables which in any given situation may or may not be significant.

REFERENCES

1. BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake, & G. V. Ramsey (Eds.), *Perception: an approach to personality.* New York: Ronald, 1951.
2. BRUNER, J. S., & POSTMAN, L. Emotional selectivity in perception and reaction. *J. Pers.,* 1947, **16**, 69–77.
3. BRUNER, J. S., & POSTMAN, L. On the perception of incongruity: a paradigm. *J. Pers.,* 1949, **18**, 206–223.
4. BRUNER, J. S., & POSTMAN, L. Perception, cognition and behavior. *J. Pers.,* 1949, **18**, 14–31.
5. ERIKSEN, C. W. Perceptual defense as a function of unacceptable needs. *J. abnorm. soc. Psychol.,* 1951, **46**, 557–564.
6. ERIKSEN, C. W. Defense against ego-threat in memory and perception. *J. abnorm. soc. Psychol.,* 1952, **47**, 230–235.
7. HOWES, D. H., & SOLOMON, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.,* 1950, **57**, 229–234.
8. HOWES, D. H., & SOLOMON, R. L. Visual duration threshold as a function of word probability. *J. exp. Psychol.,* 1951, **41**, 401–410.
9. HOWIE, D. Perceptual defense. *Psychol. Rev.,* 1952, **59**, 308–315.
10. KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948.
11. LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.,* 1951, **19**, 471–482.
12. LAZARUS, R. S., & MCCLEARY, R. A. Autonomic discrimination without awareness: a study of subception. *Psychol. Rev.,* 1951, **58**, 113–122.
13. LUCHINS, A. S. An evaluation of some current criticisms of Gestalt psychological work in perception. *Psychol. Rev.,* 1951, **58**, 69–95.
14. MCGEOCH, J. A. *Psychology of human learning.* New York: Longmans, Green, 1942.
15. MCGINNIES, E. M. Emotionality and perceptual defense. *Psychol. Rev.,* 1949, **56**, 244–251.
16. POSTMAN, L. Toward a general theory of cognition. In J. H. Rohrer, & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951.
17. POSTMAN, L. Perception, motivation and behavior. Paper read at the symposium on Cognitive Theory and Personality Functioning at the APA meeting, Washington, D. C., 1952.
18. POSTMAN, L., BRONSON, WANDA C., & GROPPER, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.,* 1953, **48**, 215–224.
19. POSTMAN, L., & BROWN, D. R. The perceptual consequences of success and failure. *J. abnorm. soc. Psychol.,* 1952, **47**, 213–221.
20. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. M. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.,* 1948, **83**, 148–153.
21. POSTMAN, L., & LEYTHAM, G. Perceptual selectivity and ambivalence of stimuli. *J. Pers.,* 1951, **18**, 390–405.
22. POSTMAN, L., & SCHNEIDER, B. H. Personal values, visual recognition, and recall. *Psychol. Rev.,* 1951, **58**, 271–284.
23. ROSENSTOCK, I. M. Perceptual aspects of repression. *J. abnorm. soc. Psychol.,* 1951, **46**, 304–315.
24. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.,* 1951, **58**, 256–270.

[MS. received September 2, 1952]

Psychological Review
Vol. 60, No. 5, 1953

RATIONAL BEHAVIOR AND ECONOMIC BEHAVIOR

GEORGE KATONA

Survey Research Center, University of Michigan

While attempts to penetrate the boundary lines between psychology and sociology have been rather frequent during the last few decades, psychologists have paid little attention to the problems with which another sister discipline, economics, is concerned. One purpose of this paper is to arouse interest among psychologists in studies of economic behavior. For that purpose it will be shown that psychological principles may be of great value in clarifying basic questions of economics and that the psychology of habit formation, of motivation, and of group belonging may profit from studies of economic behavior.

A variety of significant problems, such as those of the business cycle or inflation, of consumer saving or business investment, could be chosen for the purpose of such demonstration. This paper, however, will be concerned with the most fundamental assumption of economics, the principle of rationality. In order to clarify the problems involved in this principle, which have been neglected by contemporary psychologists, it will be necessary to contrast the most common forms of methodology used in economics with those employed in psychology and to discuss the role of empirical research in the social sciences.

Theory and Hypotheses

Economic theory represents one of the oldest and most elaborate theoretical structures in the social sciences. However, dissatisfaction with the achievements and uses of economic theory has grown considerably during the past few decades on the part of economists who are interested in what actually goes on in economic life. And yet leading sociologists and psychologists have recently declared "Economics is today, in a theoretical sense, probably the most highly elaborated, sophisticated, and refined of the disciplines dealing with action" (15, p. 28).[1]

To understand the scientific approach of economic theorists, we may divide them into two groups. Some develop an a priori system from which they deduce propositions about how people *should* act under certain assumptions. Assuming that the sole aim of businessmen is profit maximization, these theorists deduce propositions about marginal revenues and marginal costs, for example, that are not meant to be suited for testing. In developing formal logics of economic action, one of the main considerations is elegance of the deductive system, based on the law of parsimony. A wide gap separates these theorists from economic research of an empirical-statistical type which registers what they call aberrations or deviations, due to human frailty, from the norm set by theory.

A second group of economic theorists adheres to the proposition that it is the main purpose of theory to provide hypotheses that can be tested. This group acknowledges that prediction of future events represents the most stringent test of theory. They argue, however, that reality is so complex that it is necessary to begin with simplified propositions and models which are known to be unreal

[1] The quotation is from an introductory general statement signed by T. Parsons, E. A. Shils, G. W. Allport, C. Kluckhohn, H. A. Murray, R. R. Sears, R. C. Sheldon, S. A. Stouffer, and E. C. Tolman. The term "action" is meant to be synonymous with "behavior."

and not testable.[2] Basic among these propositions are the following three which traditionally have served to characterize the economic man or the rational man:

1. The principle of complete information and foresight. Economic conditions—demand, supply, prices, etc.—are not only given but also known to the rational man. This applies as well to future conditions about which there exists no uncertainty, so that rational choice can always be made. (In place of the assumption of certainty of future developments, we find nowadays more frequently the assumption that risks prevail but the probability of occurrence of different alternatives is known; this does not constitute a basic difference.)

2. The principle of complete mobility. There are no institutional or psychological factors which make it impossible, or expensive, or slow, to translate the rational choice into action.

3. The principle of pure competition. Individual action has no great influence on prices because each man's choice is independent from any other person's choice and because there are no "large" sellers or buyers. Action is the result of individual choice and is not group-determined.

Economic theory is developed first under these assumptions. The theorists then introduce changes in the assumptions so that the theory may approach reality. One such step consists, for instance, of introducing large-scale producers, monopolists, and oligopolists, another of introducing time lags, and still another of introducing uncertainty about the probability distribution of future events. The question raised in each case is this: Which of the original propositions needs to be changed, and in what way, in view of the new assumptions?

The fact that up to now the procedure of gradual approximation to reality has not been completely successful does not invalidate the method. It must also be acknowledged that propositions were frequently derived from unrealistic economic models which were susceptible to testing and stimulated empirical research. In this paper we shall point to a great drawback of this method of starting out with a simplified a priori system and making it gradually more complex and more real—by proceeding in this way one tends to lose sight of important problems and to disregard them.

The methods most commonly used in psychology may appear at first sight to be quite similar to the methods of economics which have just been described. Psychologists often start with casual observations, derive from them hypotheses, test those through more systematic observations, reformulate and revise their hypotheses accordingly, and test them again. The process of hypotheses-observations–hypotheses-observations often goes on with no end in sight. Differences from the approach of economic theory may be found in the absence in psychological research of detailed systematic elaboration prior to any observation. Also, in psychological research, findings and generalizations in one field of behavior are often considered as hypotheses in another field of behavior. Accordingly, in analyzing economic be-

[2] A variety of methods used in economic research differ, of course, from those employed by the two groups of economic theorists. Some research is motivated by dissatisfaction with the traditional economic theory; some is grounded in a systematization greatly different from traditional theory (the most important example of such systematization is national income accounting); some research is not clearly based on any theory; finally, some research has great affinity with psychological and sociological studies.

havior[3] and trying to understand rationality, psychologists can draw on (*a*) the theory of learning and thinking, (*b*) the theory of group belonging, and (*c*) the theory of motivation. This will be done in this paper.

HABITUAL BEHAVIOR AND GENUINE DECISION MAKING

In trying to give noneconomic examples of "rational calculus," economic theorists have often referred to gambling. From some textbooks one might conclude that the most rational place in the world is the Casino in Monte Carlo where odds and probabilities can be calculated exactly. In contrast, some mathematicians and psychologists have considered scientific discovery and the thought processes of scientists as the best examples of rational or intelligent behavior.[4] An inquiry about the possible contributions of psychology to the analysis of rationality may then begin with a formulation of the differences between (*a*) associative learning and habit formation and (*b*) problem solving and thinking.

The basic principle of the first form of behavior is repetition. Here the argument of Guthrie holds: "The most certain and dependable information concerning what a man will do in any situation is information concerning what he did in that situation on its last occurrence" (4, p. 228). This form of behavior depends upon the frequency of repetition as well as on its recency and on the success of past performances. The origins of habit formation have been demonstrated by experiments about learning nonsense syllables, lists of words, mazes, and conditioned responses. Habits thus formed are to some extent automatic and inflexible.

In contrast, problem-solving behavior has been characterized by the arousal of a problem or question, by deliberation that involves reorganization and "direction," by understanding of the requirements of the situation, by weighing of alternatives and taking their consequences into consideration and, finally, by choosing among alternative courses of action.[5] Scientific discovery is not the only example of such procedures; they have been demonstrated in the psychological laboratory as well as in a variety of real-life situations. Problem solving results in action which is new rather than repetitive; the actor may have never behaved in the same way before and may not have learned of any others having behaved in the same way.

Some of the above terms, defined and analyzed by psychologists, are also being used by economists in their discussion of rational behavior. In discussing, for example, a manufacturer's choice between erecting or not erecting a new factory, or raising or not raising his prices or output, reference is usually made to deliberation and to taking the consequences of alternative choices into consideration. Nevertheless, it is not justified to identify problem-solving behavior with rational behavior. From the point of view of an outside observer,

[3] The expression "economic behavior" is used in this paper to mean behavior concerning economic matters (spending, saving, investing, pricing, etc.). Some economic theorists use the expression to mean the behavior of the "economic man," that is, the behavior postulated in their theory of rationality.

[4] Reference should be made first of all to Max Wertheimer who in his book *Productive Thinking* (17) uses the terms "sensible" and "intelligent" rather than "rational." Since we are mainly interested here in deriving conclusions from the psychology of thinking, the discussion of psychological principles will be kept extremely brief (see 6 and 8, Chap. 3, 4).

[5] Cf. the following statement by a leading psychoanalyst: "Rational behavior is behavior that is effectively guided by an understanding of the situation to which one is reacting" (3, p. 16). French adds two steps that follow the choice between alternative goals, namely, commitment to a goal and commitment to a plan to reach a goal.

habitual behavior may prove to be fully rational or the most appropriate way of action under certain circumstances. All that is claimed here is that the analysis of two forms of behavior—habitual versus genuine decision making—may serve to clarify problems of rationality. We shall proceed therefore by deriving six propositions from the psychological principles. To some extent, or in certain fields of behavior, these are findings or empirical generalizations; to some extent, or in other fields of behavior, they are hypotheses.

1. Problem-solving behavior is a relatively rare occurrence. It would be incorrect to assume that everyday behavior consistently manifests such features as arousal of a problem, deliberation, or taking consequences of the action into consideration. Behavior which does not manifest these characteristics predominates in everyday life and in economic activities as well.

2. The main alternative to problem-solving behavior is not whimsical or impulsive behavior (which was considered the major example of "irrational" behavior by nineteenth century philosophers). When genuine decision making does not take place, habitual behavior is the most usual occurrence: people act as they have acted before under similar circumstances, without deliberating and choosing.

3. Problem-solving behavior is recognized most commonly as a deviation from habitual behavior. Observance of the established routine is abandoned when in driving home from my office, for example, I learn that there is a parade in town and choose a different route, instead of automatically taking the usual one. Or, to mention an example of economic behavior: Many businessmen have rules of thumb concerning the timing for reorders of merchandise; yet sometimes they decide to place new orders even though their inventories have not reached the usual level of depletion (for instance, because they anticipate price increases), or not to order merchandise even though that level has been reached (because they expect a slump in sales).

4. Strong motivational forces—stronger than those which elicit habitual behavior—must be present to call forth problem-solving behavior. Being in a "crossroad situation," facing "choice points," or perceiving that something new has occurred are typical instances in which we are motivated to deliberate and choose. Pearl Harbor and the Korean aggression are extreme examples of "new" events; economic behavior of the problem-solving type was found to have prevailed widely after these events.

5. Group belonging and group reinforcement play a substantial role in changes of behavior due to problem solving. Many people become aware of the same events at the same time; our mass media provide the same information and often the same interpretation of events to groups of people (to businessmen, trade union members, sometimes to all Americans). Changes in behavior resulting from new events may therefore occur among very many people at the same time. Some economists (for instance, Lord Keynes, see 9, p. 95) argued that consumer optimism and pessimism are unimportant because usually they will cancel out; in the light of sociopsychological principles, however, it is probable, and has been confirmed by recent surveys, that a change from optimistic to pessimistic attitudes, or vice versa, sometimes occurs among millions of people at the same time.

6. Changes in behavior due to genuine decision making will tend to be substantial and abrupt, rather than small and gradual. Typical examples of action that results from genuine decisions are cessation of purchases or buying waves, the shutting down of plants or the build-

ing of new plants, rather than an increase or decrease of production by 5 or 10 per cent.[6]

Because of the preponderance of individual psychological assumptions in classical economics and the emphasis placed on group behavior in this discussion, the change in underlying conditions which has occurred during the last century may be illustrated by a further example. It is related—the author does not know whether the story is true or fictitious—that the banking house of the Rothschilds, still in its infancy at that time, was one of the suppliers of the armies of Lord Wellington in 1815. Nathan Mayer Rothschild accompanied the armies and was present at the Battle of Waterloo. When he became convinced that Napoleon was decisively defeated, he released carrier pigeons so as to transmit the news to his associates in London and reverse the commodity position of his bank. The carrier pigeons arrived in London before the news of the victory became public knowledge. The profits thus reaped laid, according to the story, the foundation to the outstanding position of the House of Rothschild in the following decades.

The decision to embark on a new course of action because of new events was then made by one individual for his own profit. At present, news of a battle, or of change of government, or of rearmament programs, is transmitted in short order by press and radio to the public at large. Businessmen—the manufacturers or retailers of steel or clothing, for instance—usually receive the same news about changes in the price of raw materials or in demand, and often consult with each other. Belonging to the same group means being subject to similar stimuli and reinforcing one another in making decisions. Acting in the same way as other members of one's group or of a reference group have acted under similar circumstances may also occur without deliberation and choice. New action by a few manufacturers will, then, frequently or even usually not be compensated by reverse action on the part of others. Rather the direction in which the economy of an entire country moves—and often the world economy as well—will tend to be subject to the same influences.

After having indicated some of the contributions which the application of certain psychological principles to economic behavior may make, we turn to contrasting that approach with the traditional theory of rationality. Instead of referring to the formulations of nineteenth century economists, we shall quote from a modern version of the classical trend of thought. The title of a section in a recent article by Kenneth J. Arrow is "The Principle of Rationality." He describes one of the criteria of rationality as follows: "We can imagine the individual as listing, once and for all, all conceivable consequences of his actions in order of his preference for them" (1, p. 135). We are first concerned with the expression "all conceivable consequences." This expression seems to contradict the principle of selectivity of human behavior. Yet habitual behavior is highly selective since it is based on (repeated) past experience, and problem-solving behavior likewise is highly selective since reorganization is subject to a certain direction instead of consisting of trial (and error) regarding all possible avenues of action.

Secondly, Arrow appears to identify rationality with consistency in the sense of repetition of the same choice. It is part and parcel of rational behavior,

[6] Some empirical evidence supporting these six propositions in the area of economic behavior has been assembled by the Survey Research Center of the University of Michigan (see 8 and also 7).

according to Arrow, that an individual "makes the same choice each time he is confronted with the same set of alternatives" (1, p. 135).[7] Proceeding in the same way on successive occasions appears, however, a characteristic of habitual behavior. Problem-solving behavior, on the other hand, is flexible. Rationality may be said to reflect adaptability and ability to act in a new way when circumstances demand it, rather than to consist of rigid or repetitive behavior.

Thirdly, it is important to realize the differences between the concepts action, decision, and choice. It is an essential feature of the approach derived from considering problem-solving behavior that there is action without deliberate decision and choice. It then becomes one of the most important problems of research to determine under what conditions genuine decision and choice occur prior to an action. The three concepts are, however, used without differentiation in the classical theory of rationality and also, most recently, by Parsons and Shils. According to the theory of these authors, there are "five discrete choices (explicit or implicit) which every actor makes before he can act" (15, p. 78); before there is action "a decision must always be made (explicitly or implicitly, consciously or unconsciously)" (15, p. 89).

There exists, no doubt, a difference in terminology, which may be clarified by mentioning a simple case: Suppose my telephone rings; I lift the receiver with my left hand and say, "Hello." Should we then argue that I made several choices, for instance, that I decided not to lift the receiver with my right hand and not to say "Mr. Katona speaking"? According to our use of the terms decision and choice, my action was habitual and did not involve "taking consequences into consideration."[8] Parsons and Shils use the terms decision and choice in a different sense, and Arrow may use the terms "all conceivable consequences" and "same set of alternatives" in a different sense from the one employed in this paper. But the difference between the two approaches appears to be more far-reaching. By using the terminology of the authors quoted, and by constructing a theory of rational action on the basis of this terminology, fundamental problems are disregarded. If every action by definition presupposes decision making, and if the malleability of human behavior is not taken into consideration, a one-sided theory of rationality is developed and empirical research is confined to testing a theory which covers only some of the aspects of rationality.

This was the case recently in experiments devised by Mosteller and Nogee. These authors attempt to test basic assumptions of economic theory, such as the rational choice among alternatives, by placing their subjects in a gambling situation (a variation of poker dice) and compelling them to make a decision, namely, to play or not to play against the experimenter. Through their experiments the authors prove that "it is feasible to measure utility experimentally" (14, p. 403) but they do not shed light on the conditions under which rational

[7] In his recent book Arrow adds after stating that the economic man "will make the same decision each time he is faced with the same range of alternatives": "The ability to make consistent decisions is one of the symptoms of an integrated personality" (2, p. 2).

[8] If I have reason not to make known that I am at home, I may react to the ringing of the telephone by fright, indecision, and deliberation (should I lift the receiver or let the telephone ring?) instead of reacting in the habitual way. This is an example of problem-solving behavior characterized as deviating from habitual behavior. The only example of action mentioned by Parsons and Shils, "a man driving his automobile to a lake to go fishing," may be habitual or may be an instance of genuine decision making.

behavior occurs or on the inherent features of rational behavior. Experiments in which making a choice among known alternatives is prescribed do not test the realism of economic theory.

MAXIMIZATION

Up to now we have discussed only one central aspect of rationality—means rather than ends. The end of rational behavior, according to economic theory, is maximization of profits in the case of business firms and maximization of utility in the case of people in general.

A few words, first, on maximizing profits. This is usually considered the simpler case because it is widely held (*a*) that business firms are in business to make profits and (*b*) that profits, more so than utility, are a quantitative, measurable concept.

When empirical research, most commonly in the form of case studies, showed that businessmen frequently strove for many things in addition to profits or in place of profits, most theorists were content with small changes in their systems. They redefined profits so as to include long-range profits and what has been called nonpecuniary or psychic profits. Striving for security or for power was identified with striving for profits in the more distant future; purchasing goods from a high bidder who was a member of the same fraternity as the purchaser, rather than from the lowest bidder—to cite an example often used in textbooks—was thought to be maximizing of nonpecuniary profits. Dissatisfaction with this type of theory construction is rather widespread. For example, a leading theorist wrote recently:

> If *whatever* a business man does is explained by the principle of profit maximization—because he does what he likes to do, and he likes to do what maximizes the sum of his pecuniary and non-pecuniary profits—the analysis acquires the character of a system of definitions and tautologies, and loses much of its value as an explanation of reality (13, p. 526).

The same problem is encountered regarding maximization of utility. Arrow defines rational behavior as follows: ". . . among all the combinations of commodities an individual can afford, he chooses that combination which maximizes his utility or satisfaction" (1, p. 135) and speaks of the "traditional identification of rationality with maximization of some sort" (2, p. 3). An economic theorist has recently characterized this type of definition as follows:

> The statement that a person seeks to maximize utility is (in many versions) a tautology: it is impossible to conceive of an observational phenomenon that contradicts it. . . . What if the theorem is contradicted by observation: Samuelson says it would not matter much in the case of utility theory; I would say that it would not make the slightest difference. For there is a free variable in his system: the tastes of consumers. . . . Any contradiction of a theorem derived from utility theory can always be attributed to a change of tastes, rather than to an error in the postulates or logic of the theory (16, pp. 603 f.).[9]

What is the way out of this difficulty? Can psychology, and specifically the psychology of motivation, help? We may begin by characterizing the prevailing economic theory as a single-motive theory and contrast it with a theory of multiple motives. Even in case of a single decision of one individual, multiplicity of motives (or of vectors or forces in the field), some reinforcing one another and some conflicting with one another, is the rule rather than the exception. The motivational patterns prevailing among different individuals making the same decision need not be the same; the motives of the same individual who is in the same external situation at different times may likewise

[9] The quotation refers specifically to Samuelson's definition but also applies to that of Arrow.

differ. This approach opens the way (*a*) for a study of the relation of different motives to different forms of behavior and (*b*) for an investigation of changes in motives. Both problems are disregarded by postulating a single-motive theory and by restricting empirical studies to attempts to confirm or contradict that theory.

The fruitfulness of the psychological approach may be illustrated first by a brief reference to business motivation. We may rank the diverse motivational patterns of businessmen by placing the striving for high immediate profits (maximization of short-run profits, to use economic terminology; charging whatever the market can bear, to use a popular expression) at one extreme of the scale. At the other extreme we place the striving for prestige or power. In between we discern striving for security, for larger business volume, or for profits in the more distant future. Under what kinds of business conditions will motivational patterns tend to conform with the one or the other end of the scale? Preliminary studies would seem to indicate that the worse the business situation is, the more frequent is striving for high immediate profits, and the better the business situation is, the more frequent is striving for nonpecuniary goals (see 8, pp. 193–213).

Next we shall refer to one of the most important problems of consumer economics as well as of business-cycle studies, the deliberate choice between saving and spending. Suppose a college professor receives a raise in his salary or makes a few hundred extra dollars through a publication. Suppose, furthermore, that he suggests thereupon to his wife that they should buy a television set, while the wife argues that the money should be put in the bank as a reserve against a "rainy day." Whatever the final decision may be, traditional economic theory would hold that the action which gives the greater satisfaction was chosen. This way of theorizing is of little value. Under what conditions will one type of behavior (spending) and under what conditions will another type of behavior (saving) be more frequent? Psychological hypotheses according to which the strength of vectors is related to the immediacy of needs have been put to a test through nationwide surveys over the past six years.[10] On the basis of survey findings the following tentative generalization was established: Pessimism, insecurity, expectation of income declines or bad times in the near future promote saving (putting the extra money in the bank), while optimism, feeling of security, expectation of income increases, or good times promote spending (buying the television set, for instance).

Psychological hypotheses, based on a theory of motivational patterns which change with circumstances and influence behavior, thus stimulated empirical studies. These studies, in turn, yielded a better understanding of past developments and also, we may add, better predictions of forthcoming trends than did studies based on the classical theory (see footnote 10). On the other hand, when conclusions about utility or rationality were made on an a priori basis, researchers lost sight of important problems.[11]

[10] In the Surveys of Consumer Finances, conducted annually since 1946 by the Survey Research Center of the University of Michigan for the Federal Reserve Board and reported in the *Federal Reserve Bulletin*. See also 8 and a forthcoming publication of the Survey Research Center on consumer buying and inflation during 1950–52.

[11] It should not be implied that the concepts of utility and maximization are of no value for empirical research. Comparison between maximum utility as determined from the vantage point of an observer with the pattern of goals actually chosen (the "subjective maximum"), which is based on insufficient information, may be useful. Similar considerations apply to such newer concepts as "minimizing regrets" and the "minimax."

DIMINISHING UTILITY, SATURATION, AND ASPIRATION

Among the problems to which the identification of maximizing utility with rationality gave rise, the measurability of utility has been prominent. At present the position of most economists appears to be that while interpersonal comparison of several consumers' utilities is not possible, and while cardinal measures cannot be attached to the utilities of one particular consumer, ordinal ranking of the utilities of each individual can be made. It is asserted that I can always say either that I prefer A to B, or that I am indifferent to having A or B, or that I prefer B to A. The theory of indifference curves is based on this assumption.

In elaborating the theory further, it is asserted that rational behavior consists not only of preferring more of the same goods to less ($2 real wages to $1, or two packages of cigarettes to one package, for the same service performed) but also of deriving diminishing increments of satisfaction from successive units of a commodity.[12] In terms of an old textbook example, one drink of water has tremendous value to a thirsty traveler in a desert; a second, third, or fourth drink may still have some value but less and less so; an *n*th drink (which he is unable to carry along) has no value at all. A generalization derived from this principle is that the more of a commodity or the more money a person has, the smaller are his needs for that commodity or for money, and the smaller his incentives to add to what he has.

In addition to using this principle of saturation to describe the behavior of the rational man, modern economists applied it to one of the most pressing problems of contemporary American economy. Prior to World War II the American people (not counting business firms) owned about 45 billion dollars in liquid assets (currency, bank deposits, government bonds) and these funds were highly concentrated among relatively few families; most individual families held no liquid assets at all (except for small amounts of currency). By the end of the year 1945, however, the personal liquid-asset holdings had risen to about 140 billion dollars and four out of every five families owned some bank deposits or war bonds. What is the effect of this great change on spending and saving? This question has been answered by several leading economists in terms of the saturation principle presented above. "The rate of saving is . . . a diminishing function of the wealth the individual holds" (5, p. 499) because "the availability of liquid assets raises consumption generally by reducing the impulse to save."[13] More specifically: a person who owns nothing or very little will exert himself greatly to acquire some reserve funds, while a person who owns much will have much smaller incentives to save. Similarly, incentives to increase one's income are said to weaken with the amount of income. In other words, the strength of motivation is inversely correlated with the level of achievement.

In view of the lack of contact between economists and psychologists, it is hardly surprising that economists failed to see the relevance for their postulates of the extensive experimental work performed by psychologists on the problem of levels of aspiration. It is not necessary in this paper to describe these studies in detail. It may suffice to for-

[12] This principle of diminishing utility was called a "fundamental tendency of human nature" by the great nineteenth century economist, Alfred Marshall.

[13] The last quotation is from the publication of the U. S. Department of Commerce, *Survey of Current Business*, May 1950, p. 10. This quotation and several similar ones are discussed in 8, pp. 186 ff.

mulate three generalizations as established in numerous studies of goal-striving behavior (see, for example, 12):

1. Aspirations are not static, they are not established once for all time.
2. Aspirations tend to grow with achievement and decline with failure.
3. Aspirations are influenced by the performance of other members of the group to which one belongs and by that of reference groups.

From these generalizations hypotheses were derived about the influence of assets on saving which differed from the postulates of the saturation theory. This is not the place to describe the extensive empirical work undertaken to test the hypotheses. But it may be reported that the saturation theory was not confirmed; the level-of-aspiration theory likewise did not suffice to explain the findings. In addition to the variable "size of liquid-asset holdings," the studies had to consider such variables as income level, income change, and savings habits. (Holders of large liquid assets are primarily people who have saved a high proportion of their income in the past!) [14]

The necessity of studying the interaction of a great number of variables and the change of choices over time leads to doubts regarding the universal validity of a one-dimensional ordering of all alternatives. The theory of measurement of utilities remains an empty frame unless people's established preferences of *A* over *B* and of *B* over *C* provide indications about their probable future behavior. Under what conditions do people's preferences give us such clues, and under what conditions do they not? If at different times *A* and *B* are seen in different contexts—because of changed external conditions or the acquisition of new experiences—we may have to distinguish among several dimensions.

The problem may be illustrated by an analogy. Classic economic theory postulates a one-dimensional ordering of all alternatives; Gallup asserts that answers to questions of choice can always be ordered on a yes–uncertain (don't know)–no continuum; are both arguments subject to the same reservations? Specifically, if two persons give the same answer to a poll question (e.g., both say "Yes, I am for sending American troops to Europe" or "Yes, I am for the Taft-Hartley Act") may they mean different things so that their identical answers do not permit any conclusions about the similarity of their other attitudes and their behavior? Methodologically it follows from the last argument that yes-no questions need to be supplemented by open-ended questions to discern differences in people's level of information and motivation. It also follows that attitudes and preferences should be ascertained through a multi-question approach (or scaling) which serves to determine whether one or several dimensions prevail.

On Theory Construction

In attempting to summarize our conclusions about the respective merits of different scientific approaches, we might quote the conclusions of Arrow which he formulated for social science in general rather than for economics:

> To the extent that formal theoretical structures in the social sciences have not been based on the hypothesis of rational behavior, their postulates have been developed in a manner which we may term *ad hoc*. Such propositions . . . depend, of course, on the investigator's intuition and common sense (1, p. 137).

The last sentence seems strange indeed. One may argue the other way around and point out that such propositions as "the purpose of business is

[14] The empirical work was part of the economic behavior program of the Survey Research Center under the direction of the author. See (8) and also (10) and (11).

to make profits" or "the best businessman is the one who maximizes profits" are based on intuition or supposed common sense, rather than on controlled observation. The main problem raised by the quotation concerns the function of empirical research. There exists an alternative to developing an axiomatic system into a full-fledged theoretical model in advance of testing the theory through observations. Controlled observations should be based on hypotheses, and the formulation of an integrated theory need not be delayed until all observations are completed. Yet theory construction is part of the process of hypothesis-observation-revised hypothesis and prediction-observation, and systematization should rely on some empirical research. The proximate aim of scientific research is a body of empirically validated generalizations and not a theory that is valid under any and all circumstances.

The dictum that "theoretical structures in the social sciences must be based on the hypothesis of rational behavior" presupposes that it is established what rational behavior is. Yet, instead of establishing the characteristics of rational behavior a priori, we must first determine the conditions a_1, b_1, c_1 under which behavior of the type x_1, y_1, z_1 and the conditions a_2, b_2, c_2 under which behavior of the type x_2, y_2, z_2 is likely to occur. Then, if we wish, we may designate one of the forms of behavior as rational. The contributions of psychology to this process are not solely methodological; findings and principles about noneconomic behavior provide hypotheses for the study of economic behavior. Likewise, psychology can profit from the study of economic behavior because many aspects of behavior, and among them the problems of rationality, may be studied most fruitfully in the economic field.

This paper was meant to indicate some promising leads for a study of rationality, not to carry such study to its completion. Among the problems that were not considered adequately were the philosophical ones (rationality viewed as a value concept), the psychoanalytic ones (the relationships between rational and conscious, and between irrational and unconscious), and those relating to personality theory and the roots of rationality. The emphasis was placed here on the possibility and fruitfulness of studying forms of rational behavior, rather than the characteristics of *the* rational man. Motives and goals that change with and are adapted to circumstances, and the relatively rare but highly significant cases of our becoming aware of problems and attempting to solve them, were found to be related to behavior that may be called truly rational.

REFERENCES

1. ARROW, K. J. Mathematical models in the social sciences. In D. Lerner, & H. D. Lasswell (Eds.), *The policy sciences.* Stanford: Stanford Univer. Press, 1951. Pp. 129–155.
2. ARROW, K. J. *Social choice and individual values.* New York: Wiley, 1951.
3. FRENCH, T. M. *The integration of behavior.* Vol. I. Chicago: Univer. of Chicago Press, 1952.
4. GUTHRIE, E. R. *Psychology of learning.* New York: Harper, 1935.
5. HABERLER, G. *Prosperity and depression.* (3rd Ed.) Geneva: League of Nations, 1941.
6. KATONA, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.
7. KATONA, G. Psychological analysis of business decisions and expectations. *Amer. economic Rev.*, 1946, 36, 44–63.
8. KATONA, G. *Psychological analysis of economic behavior.* New York: McGraw-Hill, 1951.
9. KEYNES, J. M. *The general theory of employment, interest and money.* New York: Harcourt, Brace, 1936.

10. Klein, L. R. Assets, debts, and economic behavior. In *Studies in income and wealth*, Vol. 14. New York: National Bureau of Economic Research, 1951.

11. Klein, L. R. Estimating patterns of savings behavior from sample survey data. *Econometrica*, 1951, **19**, 438–454.

12. Lewin, K., *et al.* Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.

13. Machlup, F. Marginal analysis and empirical research. *Amer. economic Rev.*, 1946, **36**, 519–555.

14. Mosteller, F., & Nogee, P. An experimental measurement of utility. *J. political Economy*, 1951, **59**, 371–405.

15. Parsons, T., & Shils, E. A. (Eds.). *Toward a general theory of action.* Cambridge, Mass.: Harvard Univer. Press, 1951.

16. Stigler, G. J. Rev. of P. A. Samuelson's *Foundations of economic analysis*. *J. Amer. statist. Ass.*, 1948, **43**, 603–605.

17. Wertheimer, M. *Productive thinking*. New York: Harper, 1945.

[MS. received July 28, 1952]

Psychological Review
Vol. 60, No. 5, 1953

THINKING CONCEPTUALIZED IN TERMS OF INTERACTING MOMENTS [1]

PAUL McREYNOLDS

Veterans Administration, Palo Alto

One of the most striking defining features of the psychology of man is that he is a thinking creature. The ever-flowing stream of mental activity—the stream of thought, as James called it, is a most fundamental fact of our science. As Hebb (5) has pointed out, the problem of the thinking process is of pre-eminent importance, and must be dealt with if a number of subsidiary problems are to be adequately handled.

This paper presents a broad and schematic formulation concerning the thinking process. While it is highly speculative and does not touch upon a number of significant problems involved in thinking, it can be argued that all we are equipped to do in this area at the present time is to speculate—as cleverly and hopefully as we can. In a matter as important as that under consideration a broad freedom in hypothesizing appears permissible, providing that it is labeled as speculation.

Transient Functions

By *transient functions* I refer to all those mental phenomena which have a phenomenally transient existence, such as thinking, feeling, imagining, fantasying, and dreaming. Transient functions have the characteristic that they occur along a physical time dimension, so that they may be likened phenomenally to a "stream of consciousness."

We assume that underlying such mental activity is neural activity, which represents it in an invariant manner. We therefore are interested in the nature of this neural activity. We assume, on introspective grounds, that the content of the mental activity undergoes relatively constant change, and consequently we assume a relatively constant change in the underlying neural activity.

A question immediately arises: does this changing neural activity consist of changes in terms of which *specific* neural units are being discharged, or does it consist rather of changes in the *pattern* of pertinent neural activity? The first alternative implies that the discharge of a given neural unit would always indicate a given mental content; the second implies that the discharge of a given neural unit would not indicate a given mental content, but rather that its "meaning" would be a function of the pattern of neural discharges of which it was a part.

I have chosen to follow up certain implications of the second alternative. My reasons for assuming the patterning possibility are discussed elsewhere (16). Briefly, they are based upon the argument that a pattern system would be more efficient than a specific unit system for transient functions. I do not postulate any actual likeness between the neural pattern and the percept it represents. The neural pattern would be symbolic in representation and not isomorphic, topologically or otherwise.

One important implication of the pattern approach should be noted immedi-

[1] From the Veterans Administration Hospital, Palo Alto, California. This paper is a revision of a talk delivered at a Research Conference of the Bay Area Clinical Psychologists, October 4, 1951, Palo Alto. Some of the ideas presented were developed originally during discussions with Fred Attneave, who, along with Gregory Bateson and C. L. Winder, read the manuscript and made a number of helpful suggestions.

ately. If the *same* neural units may express different thought contents at different times, which is the essence of a pattern system, then these same neural units cannot also be postulated to underlie memory, i.e., those neurons whose discharge would imply a given mental content cannot be the *same* neurons which somehow maintain such of this content as may be remembered, for in the former case the neural discharge would mean different things at different times, whereas in the latter case it would be required that the neuron somehow carry relatively permanently a given meaning. The conceptualization being presented therefore supposes an anatomically, though not functionally, separate memorial system. This separation, it may be noted, is not made in Hebb's (5) theory: in his theory the neural basis of transient functions consists essentially of the discharge of those same neurons which carry the given data as memory.

The neural discharge pattern postulated to underlie transient functions may be viewed in either of two ways: first, as a succession of different discharge patterns occurring periodically or quasi-periodically; or, second, as a continuously changing discharge pattern. The first possibility implies that *changes* in the pattern would occur *between* the successive pattern discharges, whereas the second implies that the effective discharges would be continuous. It is possible that our choice from these alternatives is simply one of theoretical convenience, but it does appear that the first possibility has a certain logical advantage. If pattern discharges are to have specific meanings and effects—on behavior, memory, or consciousness—they must be specific patterns, i.e., there must be no pattern change during the discharge. This is merely another way of stating the first alternative above.

We assume, then, a succession of neural discharge patterns as underlying transient mental functions. I will refer to these successive patterns as *moments* (*M*'s). This is after the work of Stroud (20, 21), though my usage of the word is slightly different from his. The phenomenal counterpart of a *M* may be termed a *percept,* and *moment sequence* will be used to refer to any number of *M*'s which occur successively. The phenomenal counterpart on an *M* sequence would be, under appropriate conditions, a "stream of thought."

The hypothesis that thinking somehow consists of a succession of discrete "pieces" is not a new one. Bain (9, p. 245) and James (9) may be referred to in this connection. More recently a "moment function hypothesis" has been stated by Stroud (20), and McCulloch and Pitts (15) have independently developed, in a different context, what is essentially a discontinuity hypothesis. Also, Walter (24) has presented a bold theory of transient functions based largely upon a television analogy. All these persons have developed important and stimulating points of view. While my interpretation differs in many respects from theirs, I wish to acknowledge my strong indebtedness to them.

Empirical Considerations

Sufficient experimental data are not yet available to provide crucial tests of the points so far stated. There are certain findings, however, which are in accord with the hypothesis, i.e., which suggest that the brain works *as if* it were like our model, and some of these may be noted now.

Perhaps the most obvious suggestive evidence that transient mental functions occur discretely is the rhythmic, quasi-periodic nature of the EEG. There are, of course, various electrical rhythms in the brain involving complicated phase relationships and harmonics, although by no means all of the electrical activity

of the brain is periodic. Alpha is probably the best-known rhythm, though it is by no means certain that the various manifestations of 10-per-sec. activity which are termed alpha all have the same function or origin. Further, it appears that at least in certain areas alpha may be a complex with three different components (26). The suggestion here is that at least some of the alpha activity is intimately related to the postulated discrete, successive *M*'s of thought.[2]

It follows, as Walter (24) has pointed out in a different context, that the interval between *M*'s should typically approximate .1 sec. One should, therefore, expect to find in the literature certain pertinent time values of .1 sec. Several of these may be mentioned briefly. (For other examples see the sections on Voluntary Action and on Memory and Learning in this paper.)

(*a*) Dodge (4) in 1907 postulated the necessity of a clearing-up process in visual perception which precludes a succession of adequate visual fixations under .1 sec. each. (*b*) Pillsbury (27, p. 696) in 1913 concluded that the maximum rate of shift of attention is once every .1 sec. (*c*) Stein (27, p. 689) in 1928 reported that a word tends to be perceived as the same regardless of the order in which the letters are tachistoscopically presented provided the total exposure time does not exceed .1 sec. (*d*) The brightness enhancement of a flickering light is maximum when the pulsation rate is in the neighborhood of 10 per sec. (2, p. 974).

These data should not be interpreted to suggest that the brain cannot deal with stimuli or changes in stimuli of less than .1-sec. duration, as Walter (24) appears to suggest. McCulloch (15), likewise, hypothesizes .1 sec. as a scanning interval, and thus implies that the brain (at least with regard to auditory and visual stimuli) effectively handles only stimulus changes occurring every .1 sec. or more. I am not absolutely certain that I have interpreted these theorists correctly, but in any event it is to be noted that stimulus changes involving much less than .1 sec. may be reacted to. This statement is confirmed by work in stroboscopic motion, in flicker, and in flutter-fusion, in all of which changes over time values of much less than .1 sec. may influence the perception.

The situation, then, is this: certain mental phenomena appear to be based upon a mechanism with approximately a .1-sec. time unit, whereas certain other data appear to be dealt with in terms of much shorter time units. How may these generalizations be reconciled?

My suggestion is this: that we are here dealing with two different, though closely related problems. I suggest that the dynamics of data at the high level —in terms of conscious perception, awareness, decisions, changes in the train of thought, and conscious discrimination —operate in terms of the successive neurological patterns which I have referred to as *M*'s, the time unit for which approximates .1 sec. I further suggest that there are in the brain analyzing areas, the function of which is to react to changes in sensory data occurring over very brief periods before these data are expressed as *M*'s, and that the integrating time values for these areas may be much less than .1 sec. This

[2] While this interpretation assumes certain EEG waves to be an *index* of neural discharge, it *does not* assume that the observed changes in potential which constitute the EEG are due to aggregates of fiber potential spikes. Rather, it is probable that certain EEG waves represent directly a change in potential (14). A function of this might be to facilitate periodic discharges in certain neural groups. The postulated relation of certain EEG rhythms to *M* discharges might, therefore, be due to the brain waves functioning as a pacemaker.

would permit the changes per se to be represented in M's.

It may be asked: if the alpha rhythm is invoked as a rough index of successive M frequency, why is it that alpha is so easily disrupted by sensory stimulation, and what does the disruption mean? The inference that alpha refers to a quasi-periodic transient mental function appears to be contraindicated at the outset by the fact that when transient functions are, in a sense, most pronounced, as when one is giving rapt attention to a sensory stimulus, the alpha disappears.[3]

Let us examine briefly the nature of the sensory input. All, or nearly all, such information is carried on neurons, and there are three possible ways by which information may be expressed by neurons: (*a*) frequency of impulses, (*b*) frequency modulation, and (*c*) specificity of neurons, i.e., *which* neurons are involved. With respect to the first two, the data are meaningful only with regard to some time base. In order for changes in stimulus values to be dealt with accurately, these time bases should be as short as possible. The faster waves which appear when alpha disappears under sensory stimulation may reflect partially the quick, successive integration of the incoming frequency data. It would be a case not so much of alpha being blocked or masked as of its being *succeeded* by wave forms reflecting entirely different functions. This would reconcile the fact that stimulus changes occupying much less than .1 sec. can be perceived with our view that the stream of thought mechanism operates basically upon an approximately 10-per-sec. frequency.

My conjecture is that there exists, somewhere within the brain, a "center" to which all data which become part of the stream of consciousness are sent, and which "directs" our voluntary behavior. The succeeding neural discharge patterns which I have postulated to underlie the stream of thought, and which I have referred to as M's, would occur in this center, which I shall hereafter refer to as the *moment manifold*.[4] The frequency of the M's as stated earlier, is taken to approximate 10 per sec. in the normal, awake adult. When incoming sensory data are not being analyzed (in subjective terms, when no attention is being given to sensory stimuli) those areas of the brain handling such analytic functions would tend to get in harmony with the M rhythm.

With regard to the time bases of the

[3] This is the problem faced also, though in different contexts, by McCulloch and by Walter. McCulloch suggests that alpha represents, in certain occipital and temporal areas, the sweep of an alerting pulse involved in the analysis of spatial and temporal patterns, and that the reason alpha apparently disappears, i.e., is not recorded, in the presence of pertinent sensory stimulation, is that the signals caused by sensory data being scanned mask the presence of alpha. In other words, as I understand him, McCulloch believes that alpha does not *really* disappear under sensory stimulation. Walter suggests that alpha indicates, in the space receptor areas of the brain, a mechanism for continuous scanning, somewhat after the manner of television, with a frame frequency of 10 per sec. So long as the field of scan is featureless, the rhythm should be smooth and rhythmic, but in the presence of sensory stimulation of a spatial sort the succession of runs and checks would disturb this characteristic. Insofar as the preliminary analysis in the brain of spatial patterns is concerned, the hypotheses offered by these men are certainly worthy of consideration, though it is difficult to square the facts of perception with a time base as long as .1 sec.

[4] It is tempting to speculate, on the basis of the stimulating work of Penfield and Rasmussen (18) and of Thomson (22), that this "center" may be in the diencephalon. It is to be noted, however, that the hypothesis does not require a "center" in the usual anatomical sense. Proper pacing of noncontiguous discharges could be regulated by a set of neurons providing periodic nonspecific facilitation, making it possible for the M manifold to be widely dispersed.

various receiving areas for integrating incoming frequency data, I see no requirement in the present model that these different areas all have the *same* periodicities, or that such periodicities should be invariant. When a perception involves data from several different sensory modalities, however, it would seem that the different periodicities should have some fairly constant ratio to each other, and this expectation may be related speculatively to the large number of harmonics found in the brain (25, 26).

MOMENT INTERACTION

The *M* manifold may be considered in terms of input, output, and intermoment dynamics. The output consists of the periodic *M* discharges. The input, including that from the sensory areas, need not be periodic, or, if some of it is periodic, need not be at *M* frequency, or, if some of it is at *M* frequency, need not be in phase with *M* discharges.

With regard to intermoment dynamics, we postulate that the pattern of each *M* tends to persist in the *M* manifold and to affect the pattern of the next *M*. This is the same postulate I have elsewhere (16) referred to under the heading of "transient traces." It may help us to deal systematically with such matters as the continuity of thought, shifts in the sequence of thought, association of ideas, and rigidity and looseness in thinking.

There are at least two ways in which an *M* pattern might tend to persist, in the sense of affecting the next *M* discharge. (*a*) The neural pattern might tend to continue by means of recurrent circuits, and (*b*) the discharge of a given neural unit in one *M* might affect the excitability curve for the relatively immediate refiring of the same neural circuit. These possibilities are not necessarily mutually exclusive, and each of them could possibly be useful theoretically: the former in helping to explain the tendency of certain features of thought to persist, and the latter in explaining the inability of maintaining exclusive attention on a nonchanging subject for more than a very brief period.

Further conjecture along this line would not be justified here, but I believe there would be no great difficulty in designing a hypothetical model to accomplish any type of persistence that we care to postulate. The real problem is to determine exactly what we want the model to do. For the present let us postulate simply that each *M* pattern tends to persist for a brief time, and that during this time its tendency to persist gradually declines.[5] The time of persistence would have to be great enough to span to the next *M*, so that the greater the persistence, and/or the nearer in time the next *M* (*M* frequency) the greater would be the effect of the persistence in determining the make-up of the succeeding *M*. The apparent simplicity of this postulate is falsely disarming, for all manner of problems arise when one conjectures as to exactly how the persistence of M_1 and the input to the *M* manifold might combine to produce the M_2 output. Without considering these problems here, however, we postulate that the greater the persistence and/or the higher the *M* frequency, the more successive *M*'s will tend to be alike, i.e., the more

[5] The *M* manifold may figuratively be conceptualized as a mosaic undergoing relatively continuous change as a function of (*a*) aperiodic input and/or input at various periodicities, and (*b*) the disappearance of the representation of previous input. The *M*'s may then be conceptualized as periodic patterns of discharge from this mosaic. In terms of analogy, as suggested by Gregory Bateson, one may conceive of a mosaic of lights undergoing relatively continuous change. With a high-speed camera one could take pictures of this mosaic at .1-sec. intervals. These pictures would represent the discharge patterns referred to as *M*'s.

they will tend to have in common. A companion postulate is that *M* frequency and *M* persistence may vary, not only between *S*s, but also within a given *S* at different times.

In addition to *M* interaction in terms of the direct influence of one *M* on the next it is useful to postulate the existence of a mechanism whereby certain relations between two successive *M*'s could themselves be represented as input in a succeeding *M*. For example, M_3 might include some kind of representation of the extent of commonality of M_1 and M_2, or M_3 might include representation of differences in M_1 and M_2. This postulate, too, leads to all manner of difficulty when one attempts to work it out in detail, but I will not discuss those intricacies here. It, or some postulate very much like it, seems to be necessary in order to keep us from getting lost in the fallacy of conceiving of the *M* manifold as a kind of screen upon which appear successive "thoughts" with a "little man in the head" sitting and comfortably viewing the procession.[6] The *M* manifold itself is intended to *be* the "little man," and the model must provide some basis whereby the process of thought itself becomes a part of succeeding thought.

It is suggested also that *M*'s may affect the content of succeeding *M*'s by influencing, to some extent, the input to the *M* manifold from the memorial system. More specifically, it is proposed that there is at least a tendency for the *M*'s to be "filed" in the memorial system, and that in this process certain data, already filed there, become more readily available as new input to the *M* manifold. Among the factors which might determine which of the memorial data would thus be made more available would be the degree of similarity, possibly in terms of common parameters, between the input from the *M* manifold and the various memorial units. Further, I propose, as a postulate of theoretical convenience, that at any given time the extent to which the availability of given memorial data is enhanced in this manner is a function not only of a single preceding *M*, but of the additive effects of at least several preceding *M*'s, their respective influence being a function of their recency.

PERCEPTION OF SIMILARITY

Consider now the problems of conscious discrimination and the perception of the similarity of complex stimuli. In the present model all such perceptions will be considered to be based on successive comparison, i.e., to involve the comparison of successive *M*'s, in still following *M*'s, in terms of their neurological likeness. Note that I am here referring to complex stimuli, i.e., stimuli which have many dimensions. Another type of comparison of stimuli may involve what appears to be a single dimension, such as comparisons of stimuli as to brightness, pitch, hue, etc. This latter type I do not include within the present hypothesis. There is some reason to believe that it may involve an entirely different type of central mechanism, as suggested on logical grounds by James (9, ch. 13) and by Landahl (13). I think it possible that the similarity mechanism for multidimensional stimuli works on some kind of an overlap basis, and that of unidimensional stimuli on some kind of difference, or

[6] An example of this fallacy would be to consider that if successive *M*'s represented the same object just a little different in position in each *M*, as in the successive frames of a motion picture film, the phenomenal perception *ipso facto* would be motion. Such is *not* the case, even aside from the fact that the frequency of the *M*'s, as postulated, is too low to underlie stroboscopic motion. What would result would be simply successive percepts, and nothing else. Motion is itself a perception, and therefore would itself have to be represented in an *M*.

ratio, basis. While I believe that both of these could be fitted into the present model, I will restrict my present discussion to the comparison of complex stimuli.

The overlap just referred to would occur in a sequence of successive *M*'s as the comparative stimuli were being considered.[7] As to just *what* would overlap I can offer only a guess, and my guess is that it would be something like the percentage of parameters in common with the comparative stimuli. In accordance with this view is the finding by Attneave (1) of a high positive correlation between judged similarity of complex verbal stimuli and number of ways the stimuli were conceived as being alike.

The degree of overlap between successive *M*'s would be influenced by the parameters which the comparative stimuli *have* in common, and this in turn presumably would depend to some extent on the familiarity of the *S* with the comparison objects, and the attention he gives them. But it would also depend on *M* frequency and persistence.[8] (For convenience I will hereafter refer to *M* frequency as *M*f and to *M* persistence as *M*p.) In general for an *S* with high *M*f and/or marked *M*p there would tend to be more overlap between successive *M*'s than for *S*s with low *M*f and/or little *M*p.

When an *S* compares two stimuli with regard to their degree of similarity, his judgment is presumably based not only on the perceived likeness between them, but also on his adaptation level for similarity. Use of the concept of adaptation level (AL) in this context, while new, appears to be in accord with the view of Helson (6) that the concept has wide generality. Now $AL_{Sim.}$, in terms of the present model, would be some function of the central tendency of similarity represented in the multitude of past successive pairs of *M*'s. $AL_{Sim.}$ would be high (marked overlap in successive *M*'s) when *M*f is typically high and/or *M*p is typically great; the reverse would also be true. Consequently, even though a given *S*, for a given pair of comparison stimuli, might perceive the same amount of physical likenesses as another *S*, his perception of similarity, founded partly on his $AL_{Sim.}$, would tend to be a function of his typical *M*f and/or *M*p. This variable I have elsewhere (17) discussed under the name of "similarity-standards." It is to be noted that on the basis of the conceptualization proposed here this should tend to be a general trait, i.e., independent, to a large extent, of the type of stimuli being considered (so long as they are complex), since the pertinent variables are not only the particular dimensions of the stimuli, which would differ from stimulus to stimulus, but also *M*f and *M*p, which presumably are independent of the type of stimulus.

The deduction that reactivity to psychological similarity tends to be a general trait, if valid, may have significant implications for understanding individual differences in, and relationships between, all phenomena which involve psychological similarity, such as transfer, satiation and co-satiation, the substitution value of interpolated tasks, the resumption of incompleted tasks, retroactive inhibition, and the degree of heterogeneity of a train of thought.

[7] This is not to imply that comparison of percepts is typically as simple as M_1 for percept$_1$, M_2 for percept$_2$, and M_3 for their comparison, but rather that the situation cannot be simpler, and that this may be the basic paradigm of conscious perceptual discrimination. What usually happens when we compare complex stimuli is that our attention goes repeatedly from one stimulus to the other (8, p. 498), so that a whole complicated *M* sequence would be involved.

[8] The implication of "persistence" in the dynamics of psychological similarity was suggested, in a different context, by Calkins (8, p. 270) in 1892.

VOLUNTARY ACTION

One of the major types of output from the M manifold would be voluntary action. There are several areas of evidence which conform to the hypothesis that voluntary behavior is based upon a quasi-periodic function with a typical frequency of approximately 10 per sec.

1. Kibbler, Boreham, and Richter, studying the relation of the alpha rhythm and eye opening in response to an auditory stimulus, reported: "The measurements confirmed that the probability of a response was not randomly distributed in time, but showed peaks and troughs recurring approximately 10 per second and in accurate phase relations with the alpha rhythm. Similar results were obtained for other voluntary movements" (12, p. 371).

2. Reaction time has been widely studied since the last century and yet remains essentially a mystery. For example, for a single visual stimulus, the time lag between the stimulus and the most rapid response averages about .18 sec. Why should it take so long? It takes only about .01 sec. for the impulses to be transmitted to the optic nerve, and only about .01 sec. more for the motor nerve discharge to initiate movement (3).

Two men—independently, so far as I know—have formulated substantially the same hypothesis to explain this problem. Stroud (20), in his pioneering study of periodic phenomena, was struck with the discrepancy between the unpredictability of response to single isolated stimuli and the very precise timing involved in such everyday activities as piano playing and other finely coordinated rhythmic movements. Simple reaction times (RT's) are, after all, really rather astonishingly great. Reaction times of .1 to .2 sec. are, for example, of sizable value in an athletic event such as the hundred-yard dash, or in swerving one's car to prevent an accident. Stroud's hypothesis was essentially that the long RT's are a function of M periodicity. Walter (24) has suggested a very similar explanation.

This explanation can be fitted almost directly into the present model. Specifically, the explanation would be as follows. The stimulus, coming at an unpredictable time with respect to the contemporary M phase, would occur at any time between, say, plus .09 sec. to minus .01 sec. before a given M_x. This M_x would underlie awareness of the stimulus. But since there can be no change within an M, the reaction would be represented in the next M, M_y. The total time involved would thus be, for this simplest case, approximately 100 ms. plus the elapsed time before M_x, i.e., between 100 and 200 ms., the exact time being unpredictable in any given case.

Further evidence that voluntary reactions to stimuli are based upon a common periodic mechanism dealing with conscious thought processes is given by the fact that RT's to various types of stimuli tend to be highly intercorrelated (27, p. 337). That the mean RT's differ for different types of stimuli implies that different analytic mechanisms are involved in preparing the sensory data for representation in the M manifold, and that these different processes take different lengths of time for the different modalities.

3. Vince (23), in his study of tracking (reactions to continuously changing stimuli), found that *S*s attempting to make the proper movements with a control mechanism in order to follow a moving target tended to respond to the stimuli as a unit when the interval between the stimuli was less than .1 sec. This is in accord with the report of Stetson and McDill who pointed out in 1923 that, with regard to fast movements, "movement elements occurring at the rate of 10 per sec. are the units; at most then

movements can be modified not oftener than ten times in a second" (19, p. 23).

4. The maximum frequency of repetitive movements, such as tapping the finger, is approximately 10 per sec. (27, p. 696).

MEMORY AND LEARNING

Another major type of output from the *M* manifold would be that to the memorial system. The interacting *M*'s conceptualization implies that the units of information to be "filed" are the contents of the *M*'s themselves. A given unit of memory could never be perceived exactly as it was originally, however, because always there would be the persistence from the contemporary *M* sequence, and for the same reason data would never return to the memorial system unchanged.

The studies of Penfield and Rasmussen (18, pp. 179–181) suggest that memory may be filed largely in the temporal areas. The report of Kennedy and Gottsdanker (10) indicates that the kappa electroencephalogram is maximal during tasks involving active recall. Since the kappa rhythm has a frequency of 8–13 per sec. and appears to be of temporal origin, it is possible that there may be some kind of interrelation between memorial and transient functions whereby the memorial "files" are scanned, at *M*f, in what subjectively can be described as the search for a given memory.

Among the major problems of learning are these: How are separate thoughts, ideas, or perceptions linked together? How is this influenced by motivating conditions? How do concepts develop? Considering these problems in a tentative and conjectural manner, how might they be handled within the context of the interacting *M*'s model?

First, the units which would be linked together might be the *M*'s. Second, the means of linkage might be the persistence of one *M* into the next—the greater the degree of commonality between *M*'s (in learning, these would typically be successive *M*'s) the stronger the linkage.

The learning of concepts would have to be based in some manner upon the linkages, i.e., the commonalities themselves. It is as if there is some means whereby those features which sequences of *M*'s tend to have in common are filed separately, or in addition to, the *M*'s per se—and eventually can themselves become *M* manifold input. This will be recognized as very similar to James' (9, p. 506) law of dissociation by varying concomitants. The essential meaning of it here is that concepts develop from the linkages in learning, or, to put it in another way, that it is chiefly concepts which link separate perceptions together.

To bring motivation into learning in a systematic way in the present model it is necessary to postulate that degree of linkages, i.e., *M*f and/or *M*p, are influenced by motivational conditions. There is really no evidence that this is the case, but our present purpose is simply to show that it could be the case, and there are some interesting data which may be considered in this light. First, there is the frequent appearance of delta rhythms during certain emotional states.[9] This could be interpreted to suggest that fixation in learning should be less during highly emotional states. The same type of interpretation could be used to explain why dreams are so poorly remembered. There is some evidence that anxiety increases the alpha frequency, or at least makes its fast components more dominant. This would imply that linkages during anxiety should be strong. It does *not* necessarily imply that behavior learned during anxiety would tend to be voluntarily repeated, or that it would be easily recalled—only that it would be well learned. Further,

[9] The emotion itself would, of course, be represented within the *M* pattern.

there are the interesting reports by Walter (25, 26) concerning the theta rhythm (frequency 6–8 per sec.). He finds a tendency for it to appear (chiefly in the temporal area) in children with various emotions and in adults with the cessation of a pleasurable stimulus. Walter conjectures that it may be a response to "minus-pleasure." Apparently it could also be described as a pattern which may appear with frustration, as a means, perhaps, of decreasing the degree of linkage of percepts which subjectively are felt to be frustrating.

This discussion of learning and electroencephalographic findings should be interpreted as extremely tentative and speculative. The purpose has not been to prove a point of view, but simply to show that the present model could be elaborated in such a way as to include conscious learning.

Consciousness

The word *consciousness* is used here to refer not only to the capacity for selective reactivity but also to include awareness. Thus man not only has experiences, but he is aware of having experiences; he not only knows, but he knows that he knows; he not only thinks, but thinking itself can become the object of thought.[10]

Viewed in this way, the stream of consciousness is more than just a sequence of separate, independent percepts. Rather, part of the immediately past thought becomes a part of the present thought. The just past thought becomes an object of perception in the present thought, which is itself a possible object for the succeeding thought. There is continuity to the stream of consciousness, each thought giving way to the next, yet maintaining something of the previous thought. Whether this phenomenal continuity involves simply the types of interaction referred to earlier, or whether it is based upon still more obscure mechanisms, it does seem to occur.[11] In terms of the present model the extension of one "thought" into the next, by whatever mechanism, would be in terms of *M*'s, and at least two of the pertinent variables would be *M*f and *M*p.

Directly, the greater *M*f and/or *M*p, the greater the degree of consciousness and vice versa. When the succeeding *M*'s did not overlap at all, there would be no consciousness at all, and no thinking in the usual meaning of the word. This would be the case, e.g., in deep sleep or deep anesthesia. It is noteworthy that under such conditions very slow brain waves do indeed appear, though it is, of course, merely a guess that they are indicative of *M*f.

When *M*f and *M*p are such that succeeding *M*'s overlap slightly, a vague and indistinct type of consciousness should be the subjective counterpart. It would be true consciousness, however, because there could now be interaction

[10] This conception is currently out of fashion, perhaps because it seems too metaphysical. But it is a problem nonetheless. When the final understanding of the neural basis of consciousness is obtained it is unlikely that it will permit a real comprehension of just *why* that neural basis underlies consciousness. We must not expect the impossible. Such concepts as the curvature of space, a finite universe, or *n*-dimensional space cannot really be directly comprehended, but they may be useful nonetheless, because they give order and meaningfulness to the otherwise obscure. It is in this sense that we may seek the neural basis of consciousness. And when some day the solution has been found, the question of precisely *why* does this yet-to-be-understood neural system result in consciousness may be as everunanswerable a question as that of the child who asks, "*Why* does the earth pull the apple toward it?" What we may properly attempt, then, is to design a model with such characteristics as seem to be true of consciousness.

[11] My point of view here is based very largely upon James' discussion of the Pure Ego. Readers who wish a more elaborate and enlightening presentation are advised to reread his *Principles* (9, ch. 10).

of succeeding "thoughts," i.e., there would be more than the mere refiring of previously filed memory units, unaltered in any way by the process. Rather, there would be an interplay of succeeding thoughts in which the thoughts themselves could be objects of further thought. Such a state should be characterized by poor memory for the sequence of thought and a wide range of association. It might be the situation during dreams. If *M*f and *M*p are low, but not markedly so, then the state of consciousness should be more like that during drowsiness, mild toxic confusional states, anoxia, or alcoholic intoxication. Learning, discrimination, and memory should be mildly impaired. As *M*f and *M*p increase, so should the degree of consciousness. When these variables are high there should be marked homogeneity of thought sequences and pronounced fixation in learning.

SUMMARY

A speculative theory of thinking has been presented in the form of a hypothetico-neurological model. It consists essentially of the view that the stream of thought may fruitfully be conceptualized as being based upon successive neural discharge patterns. These patterns have been termed *moments* (*M*'s). This paper has examined some of the possible effects of postulated interactions within an *M* sequence with regard to the problems of similarity, voluntary action, memory, learning, and consciousness.

What is needed now are straightforward experiments crucial to the major hypotheses. At the present time a study of the perception of high-frequency meaningful stimuli is under way. It is hoped that this will yield data pertinent to the present conceptualization.

REFERENCES

1. ATTNEAVE, F. Ability to verbalize similarities among concepts and among verbal forms. *Amer. Psychologist*, 1951, 6, 270. (Abstract)
2. BARTLEY, S. The psychophysiology of vision. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
3. CRAIK, K. Theory of the human operator in control systems. II. Man as an element in a control system. *Brit. J. Psychol.*, 1948, 38, 142–148.
4. DODGE, R. An experimental study of visual fixation. *Psychol. Monogr.*, 1907, 8, No. 4 (Whole No. 35).
5. HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
6. HELSON, H. Perception. In H. Helson (Ed.), *Theoretical foundations of psychology*. New York: D. Van Nostrand, 1951. Pp. 348–385.
7. HILL, D. Psychiatry. In D. Hill, & G. Parr (Eds.), *Electroencephalography*. London: Macdonald, 1950. Chap. XI.
8. JAMES, W. *Psychology*. New York: World Publishing Co., 1948.
9. JAMES, W. *Principles of psychology*. Vol. I (reprinted). New York: Dover Publications, 1950.
10. KENNEDY, J., & GOTTSDANKER, R. The relation between the kappa electroencephalogram and recall. *Amer. Psychologist*, 1949, 4, 224. (Abstract)
11. KENNEDY, J., GOTTSDANKER, R., ARMINGTON, J., & GRAY, FLORENCE. A new electroencephalogram associated with thinking. *Science*, 1948, 108, 527–529.
12. KIBBLER, G. O., BOREHAM, J. L., & RICHTER, D. Relation of alpha rhythm of the brain to psychomotor phenomena. *Nature*, 1949, 164, 371.
13. LANDAHL, H. Neural mechanisms for the concepts of difference and similarity. *Bull. math. Biophysics*, 1945, 7, 83–88.
14. LI, C., MCLENNAN, H., & JASPER, H. Brain waves and unit discharge in the cerebral cortex. *Science*, 1952, 116, 656–657.
15. MCCULLOCH, W. S., & PITTS, W. How we know universals. The perception of auditory and visual forms. *Bull. math. Biophysics*, 1947, 9, 127–147.
16. MCREYNOLDS, P. Logical relationships between memorial and transient functions. *Psychol. Rev.*, 1950, 57, 140–144.
17. MCREYNOLDS, P. The similarity-standards hypothesis. *Amer. Psychologist*, 1951, 6, 496. (Abstract)

18. PENFIELD, W., & RASMUSSEN, T. *The cerebral cortex of man.* New York: Macmillan, 1950.

19. STETSON, R., & McDILL, J. Mechanisms of the different types of movement. *Psychol. Monogr.*, 1923, **32**, No. 3 (Whole No. 145). Pp. 18–40.

20. STROUD, J. The moment function hypothesis. Unpublished master's thesis Stanford Univer., 1948.

21. STROUD, J. The psychological moment in perception. In Van Foerster (Ed.), *Conference on cybernetics, transactions of the sixth conference.* New York: Josiah Macy, Jr. Foundation, 1949. Pp. 27–63.

22. THOMSON, G. Cerebral area essential to consciousness. *Bull. Los Angeles neurol. Soc.*, 1951, **16**, 311–334.

23. VINCE, M. A. The intermittency of control movements and the psychological refractory period. *Brit. J. Psychol.*, 1948, **38**, 149–157.

24. WALTER, G. Electroencephalography. In G. W. T. H. Fleming (Ed.), *Recent progress in psychiatry.* Vol. II. London: J. and A. Churchill, Ltd., 1950.

25. WALTER, G. Normal rhythms—their development, distribution, and significance. In D. Hill, & C. Parr (Eds.), *Electroencephalography.* London: MacDonald, 1950. Chap. VII.

26. WALTER, G. The twenty-fourth Maudsley lecture: The functions of electrical rhythms in the brain. *J. ment. Sci.*, 1950, **96**, 1–31.

27. WOODWORTH, R. *Experimental psychology.* New York: Henry Holt, 1938.

[MS. received August 28, 1952]

Psychological Review
Vol. 60, No. 5, 1953

CLASSICAL CONDITIONING AND HUMAN WATCH-KEEPING [1]

D. E. BROADBENT

Applied Psychology Research Unit, Cambridge, England

It is the main thesis of the present paper that classical conditioning derives its chief interest and importance from the fact that it reduces learning to a much less central position than any other form of animal experiment. Paradoxical though such a position may seem, it will be shown that it enables us to include in a single category experimental results drawn both from conditioning and from work on human performance, the latter in situations normally regarded as extremely remote from the work of Pavlov. It is not, of course, intended to argue that learning does not appear in conditioning or even that Pavlovian types of experiment may not reveal useful information about the circumstances under which it will take place. But many of the facts of conditioning are not necessarily related to learning, and, insofar as existing theories of learning attempt to include them, they may well be engaging in an impossible task. The following remarks will not add still another theory of learning to the existing array, but rather try to clear the ground for future theorizing in this area of behavior.

Historically, psychologists have tended to regard conditioning primarily as a case of learning. Yet Pavlov himself regarded his own work as intimately related to the investigation of spinal reflexes by Sherrington (28, p. 378). Such a view gave importance not only to learning but also to further relations between learned responses once acquired; there was no learning among the Sherringtonian properties of the reflex, but there was an abundance of such phenomena as spatial and temporal summation, inhibition of one reflex by another, reciprocal innervation, and similar transitory rather than permanent alterations in response. Equally Pavlov sought to establish principles which would favor or impede the appearance of responses on particular occasions.

The physiological approach, however, caused Pavlov to assume that all stimuli falling on the receptor field produce some tendency for a response to appear, so that interaction between different responses takes place at the final common path as in spinal reflexes. Such an assumption may not be valid, and it is worth considering the consequences of denying it. It can scarcely be claimed that the individual ideas set forth in what follows are original; besides those which are given specific references, the general principles asserted are to be found in a different language in a celebrated chapter by William James (19, ch. XI)—a chapter which opens with a denunciation of the emphasis placed by his comtemporaries on learning alone! Some degree of novelty, however, may be found in the application of these principles to classical conditioning.

A Fresh Analysis of the Pavlovian Situation

There is some mathematical evidence to suggest that even the human nervous system would be incapable, through limitations of size, of simultaneously analyzing all the information received by its sense organs (16). It is not unreasonable, then, to suppose, that only a

[1] This paper was written while the author was receiving a grant from the British Medical Research Council. What is good in it is probably due to Professor Sir Frederic Bartlett, F.R.S., and Dr. N. H. Mackworth.

portion of this information is analyzed at one time, that is, that only one part of the total stimulation present is capable of initiating complex responses at a given instant. This suggestion has been linked by Hick to the subjective concept of "attention" (16). The possibility is further favored by the long succession of experiments on perceptual set in human beings, some of which have shown that adequate response to one part of the stimulus situation is incompatible with adequate response to another part. This has recently been demonstrated for the auditory field, thus eliminating questions of foveal and peripheral sensitivity (11). Such incompatibility may perhaps appear only in adults and be developed by learning from an original unselected reception of information, but it should be remembered that Pavlovian conditioning is normally carried out on adult mammals of reasonably high complexity. A process of selection among the various stimuli presented is, therefore, not unlikely to appear in such experiments.

There is some evidence, of which the references to be cited are representative rather than exhaustive, that certain properties of a stimulus will make it more likely to be selected from among its competitors, and to retain its dominance if selected by chance. *Physical intensity* is one such property (2, 12). Another is position of the general class of similar stimuli in a hierarchy which is independent of physical intensity but may be established by experiment (e.g., stimulation of a pain receptor might have high priority, and so might food for an animal which has been deprived of it). This property may be called for brevity *biological importance*, and as the above examples show, it may be established innately or by learning, and may be permanent or dependent on the temporary state of the organism (21, 34). A third property of the stimulus is likely to be important for selection, unless the organism is to be dominated continuously by one stimulus; this is degree of difference from the preceding or more remotely past selected stimulus, that is, *novelty* (1, 4). There is no need to assume that the selected aspects of the situation should produce any overt response, but if a stimulus is to produce a response it must first be selected. Appearance of a response which is known to depend on a certain stimulus may then inform us that this stimulus is dominant, but absence of response may indicate merely that the dominant stimulus has no dependent response. In this case, as will be seen, responses to subsequent and simultaneous stimulation may be affected.

Such principles do not involve learning in the sense of a permanent change in the response produced by a particular stimulus; yet they are capable of accounting for many of the features of conditioning.

(*a*) If the stimulus which an experimenter applies is to produce any response, innate or acquired, it must possess marked advantages over all competitors: The animal's surroundings must therefore remain unchanged for a considerable time before the experiment, contain no important objects such as humans, and not contain stimuli of high physical intensity (28, p. 20).

(*b*) If a stimulus is repeatedly applied, as its novelty becomes less it is less likely to be selected and so to produce any response. A time interval after such a repeated application will permit selection of other stimuli and so, by the third of the above principles, once more give a preference to the original stimulus. (The familiar phenomenon of extinction and spontaneous recovery. It will be noted that this is a theory of extinction by "competing stimulus." Its advantages over existing theories will be discussed later.)

(*c*) If between two presentations of the original stimulus there is inserted a different stimulus having a high priority according to these principles, the latter will be selected; when the original stimulus is next presented, it will then be more markedly different from the preceding selected stimulus, and so by the third principle will be more likely to produce a response. (Disinhibition, and also the retarding of temporary extinction by continued reinforcement.)

(*d*) The foregoing result will naturally apply only if there is a moderate interval between the two stimuli. If the fresh high priority stimulus is presented simultaneously with, or very closely previous to, the original stimulus, it will markedly lower the probability of the latter being selected. (External inhibition, and the difficulty of backward conditioning.)

(*e*) If a series of presentations of the original stimulus is followed by the presentation of a different and moderately high priority stimulus to which the acquired response is inaction, this will have an effect similar to that of disinhibition. ("Positive induction" [28, p. 189]. This is a case where a stimulus having no response has a subsequent effect.)

(*f*) This induction effect will presumably vary with time since the presentation of the intruding stimulus and with the similarity of the two stimuli. ("Irradiation" [28, pp. 153 f.]. It is not necessary to suppose that the function takes the particular form stated by Pavlov, but only that both variables enter into it [22].)

(*g*) The presentation of a reinforcing stimulus only after every other or by every third presentation of the conditioned stimulus may enhance its effect by preventing reinforcement from losing its "novelty" priority. ("Partial reinforcement" [28, p. 334]. This effect will naturally be most marked when trials are massed [20].)

(*h*) Extending the time of presentation of a stimulus is likely to change the relative priorities of different stimuli, a formerly high priority stimulus becoming low and vice versa. ("Paradoxical phase" [28, pp. 379 f.].)

(*i*) Similar effects will appear not only with learned responses but also with unlearned ones: thus the lid response to moderate sounds will disappear with repeated presentation and reappear when a different stimulus is inserted between presentations. Increase in the intensity of the stimulus or use of a stimulus of greater biological importance makes such a disappearance unlikely (27).

Some comments may be made on these points. It will be noted that the two principal novel concepts are the classing of reinforcement with disinhibition and the use of what may be called a "competing stimulus" theory of extinction as opposed to the existing theories of competing response or of depression of activity (17, pp. 115 f.).

As regards the role of reinforcement, it should be emphasized once again that it is not intended to advance a theory of learning. Rewards may or may not be necessary to produce a change in the response produced by a given stimulus; all that is being argued is that after learning has taken place the stimulus may be made to secure repeated selection by interspersing presentations with other stimuli of high priority—and that in this respect novel or intense stimuli should be classed together with those of biological importance. Such a view has advantages in considering the factors which maintain an established type of response in human beings and also in some experiments showing that "punishing" stimuli as well as "rewarding" ones (26) may assist performance. Further, it is supported by the work of Pavlov

(28, p. 390) showing that reinforcement of an extinguished response has an effect as transitory as disinhibition. Contiguity theorists may perhaps feel that the present approach offers a promising line of assault on existing reinforcement theories, particularly when recent work on reinforcement without drive reduction is considered (32, 33), but they will be going beyond the evidence in doing so.

The question of the general relationship of this approach to learning theory also arises in regard to the "continuity" controversy: Is all stimulation presented to an organism equally likely to form learned connections to responses? This question again lies outside the present scope; it is generally admitted that in the adult, selective response does appear, but it may be true to say that it is a consequence of earlier learning or that learning of other stimuli may be going on though undetected. The present argument is, however, that the facts of extinction follow from those of selective response. Consequently, it seems illogical to neglect the latter in learning theory and yet to include the former. If either is to be included, it should be selective response since it is logically prior. It should also be noted that existing tests of continuity are inadequate: they only contrast unselected learning with completely selective learning. It is also possible (using the language of the present theory) that only selected stimuli are learned, but that under favorable circumstances the selector mechanism will "scan" different stimuli in succession. Spaced learning might, on such a view, favor continuity theories, while massing would oppose them: this is in fact the case (6). But existing data are insufficient to shed light on this point.

We cannot leave the question of learning theory without noting the capable and painstaking suggestions of Berlyne (1, 2, 3, 4) to link phenomena of the sort considered here with the Hullian system. His emphasis on an internal "perceptual response" and his classing of novelty as a "drive" appear to be similar in many ways to the present view, except that he applies the same Hullian postulates both to the external responses for which they were first stated and to his internal one, whereas the writer would suspect that different principles will appear in the two cases and that the system of Hull confuses the two. Most probably this divergence reflects a general difference of opinion on the relative value of theories which cover all behavior but which require numerous simultaneous assumptions and of specific theories whose scope is limited but whose ratio of assumptions to observables is rather lower. On this point the writer's view of scientific method differs very considerably from that of Hull's supporters. It is interesting to note that Hull himself mentioned some of the possibilities suggested here, in particular to explain the difficulty of delayed conditioning (18, pp. 207 f.), but apparently did not realize that they were alternative, rather than additional, to an orthodox extinction theory.

As for the advantages of a theory of extinction by competing stimulus, it will be seen that it avoids the weaknesses of both the existing theories. On the one hand, it would be expected that extinction would be favored by massing rather than spacing while acquisition would reverse this relation, and that correlations between rates of acquisition and extinction would not be positive. These points are usually regarded as favoring extinction by depression of activity; but the detectable presence of responses to competing stimuli and the hastening of extinction when these responses are encouraged tells against this last view while being perfectly consistent with a theory of extinction by competing stimulus. It is, of course, only tempo-

rary extinction that is covered by this theory: permanent extinction is a problem of learning to refrain from response and so falls into the field of learning theory. A two-factor theory of inhibition is already current (18).

The explanation of Pavlov's less central concepts must be regarded as tentative, since accounts of the experiments seem to indicate that the name "paradoxical phase," for example, was used for several states arising from different procedures and possibly different in nature. From the present standpoint, it would seem that Pavlov's terminology was led astray by his belief that interference between different stimulus-response links took place at the level of the motor pathway. He could not, for example, assimilate positive induction to disinhibition because the latter required, for Pavlov, an active response to inhibit previous inhibition, and the negative inducing stimulus had no such active response (28, p. 191). The illustrations given show, however, that the more exotic features of Pavlovian conditioning are open to reasonable explanation from the present viewpoint and deserve a more thorough experimental investigation than they have usually received. We reach here the central peculiarity of classical conditioning—that the dominance of the situation by the experimentally controllable stimulus makes it possible for us to follow the movements of the animal's attention, using the learned response as a detecting instrument. The principles of attention are no doubt operative in all behavior, including instrumental conditioning and other laboratory situations, but they are always confused with learning and with other factors in a manner hardly resolvable except by instructing the animal on the requirements of the task—and that can be done only with one species. As soon as emitted responses are involved, moreover, we face the problem of the relation between the perceptual mechanism and that which controls such emitted responses, for it is not clear how far the two are independent, or, in common speech, how much we must attend to what we do (11). In such a field our problems must be separated for experiment, so far as that is possible without artificiality, and for the study of the processes which select now one part of the environment for response, now another, Pavlov's technique has no rival in infrahuman animals. His goal was to introduce order into the apparently lawless behavior of organisms, which, faced with the same situation, do not always react in the same way. His analysis of his results (and especially his speculative physiology) may have been imperfect, but his experiments were far more subtle than his successors have usually believed.

PAVLOV AND APPLIED PSYCHOLOGY

The most critical of the deductions made in the previous section is probably the last. Sharp lines divide most theories of learning from reflex behavior on the one hand and from the performance of the human worker on the other, lines bridged only by a belief in the unity of all behavior. Reference has already been made, however, to the data given by Oldfield (27) on the lid responses to sound; similar observations given in the same paper are equally understandable on the present view. We may perhaps note here in particular, as they have been criticized, the results drawn from Rawdon-Smith (29), who found that the auditory threshold was elevated after exposure to a high intensity tone, that this applied to the opposite ear as well as that exposed, and that the elevation was temporarily abolished by a novel visual stimulus. Some of these results were not borne out by later workers (e.g., Caussé and Chavasse [13]). Rawdon-Smith, however, gave

his tests (to an *S* in a featureless room) for a prolonged period. His stimulating tone lasted two to four minutes while most of the stimuli employed by Caussé and Chavasse were of half a minute or less, and the period of threshold testing continued for some time after exposure —more than five minutes in the experiments on "disinhibition." The later workers used a shorter testing technique. It would be expected on the present view that such prolonged testing would cause some elevation of the auditory threshold even if no high intensity tone was applied during the series and that elevation due to this cause would be abolished by a novel stimulus. This is a normal hazard in audiometry; the writer has often observed similar elevations while giving lengthy audiograms to naive *S*s, making it necessary to introduce variations in the procedure if the lowest possible threshold is to be obtained for each *S*. Again, the operating instructions for one commercial audiometer include the following statement: "Do not keep the tone interrupted for long periods, as the patient will tire, due to the strain of listening: conversely do not keep the tone on for excessive periods—not only may the patient's attention wander, but also he may tire of holding in the signalling button" (24, p. 7). Those who fail to find elevation of the threshold on the unstimulated ear may, then, have avoided this danger as a normal point of technique, and Rawdon-Smith's results can be explained on the present view.

This explanation is more satisfactory than that given by Caussé and Chavasse themselves, who feel that Rawdon-Smith may have been detecting an elevation due to sound leaking from one ear round to the other. But Rawdon-Smith had failed to detect any elevation *using his testing technique* when applying a stimulating tone of the intensity to be expected from leakage. The work of Caussé and Chavasse is highly important simply because it does eliminate central factors of this general, not peculiarly auditory, sort, and so sheds more light on peripheral auditory fatigue and thus on theory of hearing: but this does not mean that the central factors are unreal.

The most direct analogy to Pavlovian conditioning, however, lies in tasks where the human *S* is required to make a voluntary response whenever an infrequent signal occurs. We would expect that, despite the almost immediate learning displayed in such a situation, the characteristic features of extinction would appear. The first statement of this parallelism was made by Mackworth (23) for performance on the Clock Test, a task similar to radar watch-keeping.

It appeared that:

1. Responses to the infrequent stimulus became much less common after a period of continuous watch. (Extinction.)

2. A short interval between periods of testing temporarily restored full efficiency. (Spontaneous recovery.)

3. A novel stimulus (a telephone message) temporarily restored full efficiency. (Disinhibition.)

4. The following of every stimulus by an auditory message informing *S* of his performance prevented deterioration. (Continued reinforcement.)

To these similarities others have been added by Broadbent (7), using a dial-watching task.

5. When signals come from several possible sources, some of these sources are naturally more likely to receive a response than others, but when the time of monotonous watch is extended these priorities are reversed and the initially neglected sources become favored. (Paradoxical phase.)

6. On the individual *S*s, there is a correlation between degree of decrement in performance during each watch and amount of recovery in the interval between watches. The extremes of this

correlation correspond to Pavlov's description of excitable and inhibitable types (28, p. 285).

7. The signal itself introduces a novelty into the situation, and the occurrence of each signal restores performance partially to that appearing at the beginning of the watch. (Disinhibition by a novel positive stimulus [28, p. 62].) This finding allows us to note a further point in the work of Mackworth (23), namely that *S*s who respond to more signals maintain performance better.

We also note that:

8. On the Clock Test and radar test the effect of intervals between signals was complicated by "expectancy" effects, the intervals being shorter than those used by Broadbent. One of these effects was that a rapid series of signals actually resulted in a lowered probability of response, since, subjectively speaking, the *S* expected long intervals between signals and so relaxed his attention after each. This is comparable to one of the Pavlovian demonstrations of "negative induction" (28, p. 202).

There seem, therefore, to be a number of observational analogies between the conditioning situation and that of human watch-keeping, and the later work has supported Mackworth in suggesting these analogies. Such a similarity makes an identity of underlying mechanism very probable. The present writer departs from the original discoverer of the parallelism, however, in interpreting conditioning in the light of attention rather than interpreting vigilance experiments in the language of conditioning. Certain results may be noted from experiments on human workers in other situations which, while not directly comparable to conditioning, confirm the general validity of the principles stated.

1. If a task is set to an *S* and his responses to it are allowed to appear at their natural rate, the task will after a while cease to be novel and so stimuli from the remainder of the environment will tend temporarily to weaken its dominance over response. Any such weakening will immediately render the task more novel, leading to its resumption, and so performance will not so much deteriorate as become irregular in time (5, 8). Removing the timing of the task from *S*'s control will cause an actual decrement in output with continued performance (8). (The relation between this situation and the watch-keeping one is considered in greater detail in the latter paper.)

2. The presence of an intense but irrelevant stimulus, such as noise, as well as the task to be performed, will initially have little effect on the task. There may be a very transient decrement due to an "external inhibition" effect, but this will pass (14). Indeed it may be succeeded by actually improved performance by the "disinhibition" mechanism (25, 31). As the task is continued for prolonged periods, however, the strong competing stimulus will accentuate any tendency to show a decrement (8, 9). Serious effects of noise will therefore only appear on prolonged tasks and especially on those which require continuous reception of information from a particular part of the surroundings.

3. It has been pointed out by Gagné (15) that a number of studies carried out at the USAF Perceptual and Motor Skills Research Laboratory on the learning of motor skills under massed and distributed practice are difficult to interpret on the basis of the classical "inhibition" view, but rather suggest the importance of variations in the stimulus complex.

It will be remembered, then, that originally we stated three principles as governing the selection of a stimulus for further analysis and response from among those present. These principles

account adequately for a great part of the data from conditioning and for some of that on continuous human performance where learning is reduced considerably. They are not principles in the sense of postulated states within the organism, whether excitatory, inhibitory, satisfactory, or expectant, but rather in the sense of generalized rules about the properties of stimulation important for certain behavior to appear. Equally they are not concerned with permanent changes in behavior, but with transitory and conditional ones, and it will be felt a sufficient advance if they are accepted as such. A theory of learning is an ultimate aim for psychology as a whole, but we still have no adequate evidence to sustain one.

Finally, it is felt that this approach reduces linguistic questions to a proper subsidiary position. Some may prefer to speak of the extinction of responses by human *S*s in vigilance tests, others of the boredom of Pavlov's dogs, and still others of a shift in the source of the information flowing through the receptor field. There are no doubt reasons for preferring each such language, but the important point is the similarity of the experimental relations described in them. Further experiments, too, are badly needed. As examples we may suggest:

(*a*) Attempts to condition responses to stimuli which have been presented many times while still indifferent. (The present theory would suggest that this would be very difficult, and that this might explain the difficulty of demonstrating sensory preconditioning [30].)

(*b*) Determinations of the optimum time interval between signals of different kinds and in various modalities, when such signals are used for a lifelike skilled task, or in animals, to elicit conditioned responses.

(*c*) Closer examination of the influence of background stimuli, both relevant and irrelevant, upon the execution of skills, and particularly of the effects of similarity between such stimuli and those central to the skill, and of the time relationships involved.

Such experiments must be welcomed whether they are described as dealing with induction and irradiation, with the influence of dissimilarity and sequence in time on attention, or with the corresponding informational analysis of the same problem.

SUMMARY

1. Evidence both from mathematical considerations and from experiment favors the view that only certain aspects of the total stimulus situation can initiate complex responses at one time, and stimuli possessing intensity, biological importance, and novelty are most likely to be selected at any time.

2. These principles enable us to account for many phenomena of classical conditioning, including some which are usually neglected, and give an explanation of extinction (in terms of competing stimuli rather than competing responses) which has advantages over existing theories.

3. They further provide a rational interpretation of certain experiments on human beings which are usually regarded as being of purely industrial importance; for example, the pacing of factory work by the speed of the machine may prevent occasional shifts in selection and so by the novelty principle increase apparent fatigue.

4. They do not include an account of learning, as a permanent change in response to a particular stimulus, but are intended to remove certain confusing phenomena from the field to make way for future theories of learning. It is suggested that Pavlovian conditioning is of importance as making an account of these phenomena possible and that further experiment on such matters as

Pavlovian "induction" is highly desirable.

REFERENCES

1. BERLYNE, D. E. Novelty and curiosity as determinants of exploratory behaviour. *Brit. J. Psychol.*, 1950, **41**, 68–80.
2. BERLYNE, D. E. Stimulus intensity and attention in relation to learning theory. *Quart. J. exp. Psychol.*, 1950, **2**, 71–75.
3. BERLYNE, D. E. Attention, perception, and behavior theory. *Psychol. Rev.*, 1951, **58**, 137–146.
4. BERLYNE, D. E. Attention to change. *Brit. J. Psychol.*, 1951, **43**, 269–278.
5. BILLS, A. G. Blocking: a new principle in mental fatigue. *Amer. J. Psychol.*, 1931, **43**, 230–245.
6. BLUM, R. A., & BLUM, J. A. Factual issues in the continuity controversy. *Psychol. Rev.*, 1944, **56**, 33–50.
7. BROADBENT, D. E. The twenty dials test under quiet conditions. Cambridge: Appl. Psychol. Unit Rep. No. 130, 1950.
8. BROADBENT, D. E. Noise, paced performance and vigilance tasks. Cambridge: Appl. Psychol. Unit Rep. No. 165, 1951.
9. BROADBENT, D. E. The twenty dials and twenty lights test under noisy conditions. Cambridge: Appl. Psychol. Unit Rep. No. 160, 1951.
10. BROADBENT, D. E. Listening to one of two synchronous messages. *J. exp. Psychol.*, 1952, **44**, 51–55.
11. BROADBENT, D. E. Speaking and listening simultaneously. *J. exp. Psychol.*, 1952, **43**, 267–273.
12. BROWN, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209–226.
13. CAUSSÉ, R., & CHAVASSE, P. Études sur la fatigue auditive. *Anné Psychol.*, 1947, **43–44**, 265–298.
14. FORD, A. Attention-automatization: an investigation of the transitional nature of mind. *Amer. J. Psychol.*, 1929, **41**, 1–32.
15. GAGNÉ, R. A. Work decrement in the learning and retention of motor skills. In *Ergonomics:* "Symposium on Fatigue." London: H. K. Lewis, in press.
16. HICKS, W. E. Some studies of human skill and their implication for a theory of brain mechanisms. Unpublished doctor's dissertation for M.D., Univer. of Durham, England, 1949.
17. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
18. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
19. JAMES, W. *Principles of psychology.* New York: Holt, 1890.
20. JENKINS, W. O. & STANLEY, J. C. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
21. LEVINE, R., CHEIN, I., & MURPHY, G. The relation of the intensity of a need to the amount of perceptual distortion: a preliminary report. *J. Psychol.*, 1942, **13**, 283–293.
22. LOUCKS, R. B. An appraisal of Pavlov's systematization of behavior from the experimental standpoint. *J. comp. Psychol.*, 1934, **18**, 305–313.
23. MACKWORTH, N. H. *Researches in the measurement of human performance.* M.R.C. Special Report No. 268. London: H.M.S.O.
24. MARCONI INSTRUMENTS, LTD. Operating instructions No. E.B. 895 Pure Tone Audiometer Type T.F. 895, p. 7. Section 2–4. St. Albans, 1949.
25. MORGAN, J. J. B. The overcoming of distraction and other resistances. *Arch. Psychol.*, 1916, **5**, No. 35.
26. MUENZINGER, K. F. Motivation in learning. I. Electric shock for correct response in the visual decrement habit. *J. comp. Psychol.*, 1934, **17**, 267–277.
27. OLDFIELD, R. C. Some recent experiments bearing on internal inhibition. *Brit. J. Psychol.*, 1932, **28**, 28–42.
28. PAVLOV, I. P. *Conditioned reflexes.* London: Oxford Univer. Press, 1927.
29. RAWDON-SMITH, A. F. Experimental deafness: further data upon the phenomena of so-called auditory fatigue. *Brit. J. Psychol.*, 1936, **26**, 233–244.
30. REID, R. L. A test of sensory pre-conditioning in pigeons. *Quart. J. exp. Psychol.*, 1952, **4**, 49–56.
31. RYAN, T. A. Interrelations of the sensory systems in perception. *Psychol. Bull.*, 1940, **37**, 659–698.
32. SHEFFIELD, F. D., & ROBY, T. B. Reward value of a non-nutritive sweet taste. *J. comp. physiol. Psychol.*, 1950, **43**, 471–481.
33. SHEFFIELD, F. D., WOLFF, J. J., & BACKER, J. Reward value of copulation without sex drive reduction. *J. comp. physiol. Psychol.*, 1951, **44**, 3–8.
34. YOUNG, P. T. Relative food preferences of the white rat. *J. comp. Psychol.*, 1932, **14**, 297–319.

[MS. received September 26, 1952]

Psychological Review
Vol. 60, No. 5, 1953

THEORY CONSTRUCTION FOR BLODGETT'S LATENT LEARNING

JOSEPH WOLPE
University of the Witwatersrand, Johannesburg

Comments on the Hullian Approach

When a Hullian insists that his theoretical constructs shall be derived *only* from observations of stimulus and response he handicaps himself in two ways. (*a*) He cuts himself off from helpful neurological information; for when behavioral facts can be related to each other in several theoretically possible ways the neurological facts may point to the exclusion of certain of the possibilities (21, 22, 23, 24, 25). For instance, neurological facts were shown (25) to contradict Spence's prediction of the character of the gradient of generalization of extinction effects (15), when on the available behavioral facts alone that prediction was just as plausible as any other. (*b*) He loses sight of the fact that the events that intervene between stimulus and response take place *inside the organism*, and that consequently behavioral antecedents may have intraorganismal consequents *without behavioral consequents.*

It is their failure to realize the latter possibility that has made latent learning such a stumbling block to Hullian theorists. Latent learning provides no behavioral consequent to which a Hullian construct can be anchored. Therefore, committed to the assumption that there can be an all-inclusive theory of behavior in terms of their characteristic constructs, the Hullians have no real answer when their inability to account for latent learning is pointed out by such critics as Hilgard (5), Thistlethwaite (16), and Birch & Bitterman (1). Consequently, some Hullian theorists have been understandably anxious to explain away latent learning. For instance, Miller (10) has taken the view that latent learning is merely a case of very weak learning. Nobody denies that a weak measure of manifest learning accompanies latent learning, but what has to be explained is the remarkable improvement in manifest learning immediately after the introduction of food reward in the experiments of Blodgett (2) and Tolman and Honzik (19). Maltzman (8) has ingeniously tried to explain the improvement on the basis that during the unrewarded trials there is, through a process of habituation, a gradual weakening of responses in competition with the response finally to be rewarded. But, as Thistlethwaite (17) has pointed out, Herb's observations (4) contradict this proposition, and in any case experimental experience in general does not seem to show that habituation to apparatus accelerates conditioning to such an inordinate extent.

Comments on the Tolman Approach

The critics of the Hullian position have, by contrast, been quite happy about latent learning, explaining it by hypotheses of the Tolman pattern. In theorizing thus their advantage is that they are not merely correlating molar facts of behavior, but are also, in a way, *looking inside the organism.* The inner events they postulate are "expectancies" and "cognitive maps." Their explanation of the dramatic improvement in performance following the introduction of food in the Blodgett type of experiment is this: The feeding changes the animal's goal-box expectancy so that on his next run his hunger drive motivates him to use his "cognitive map" of the

maze in such a way that he takes the path to the goal box rather than other paths.

Although in this explanation such words as "expectancy" are objectively[1] defined (18) the definitions have a looseness that seems to stem from the mentalistic tradition of the words. We shall examine below whether the explanation can still be maintained if what Tolman means by "expectancy" is expressed in more rigorous terms.

It may be said that an expectancy is formed when the stimuli S_1 and S_2 have been presented to an organism in temporal contiguity a number of times, and it is subsequently found that presentation of S_1 alone causes a cognition of S_2 to be evoked in the organism. The expectancy is, then, the evocation by S_1 of the cognition of S_2.

Now let us examine more closely the mechanism by which this expectancy is evoked. It is self-evident that this mechanism must consist of internal organismal *responses* to S_1. When, for example, in an experiment on sensory preconditioning (e.g., 3) the repeated co-occurrence of stimuli S_1 and S_2 enables S_1 alone to evoke responses characteristic of S_2, these responses could well include such responses as would give the organism a *cognition* of S_2—this, by definition, constituting an expectancy. But as this expectancy would merely be a consequent of some of the S_2-connected responses it could certainly not be invoked to explain how S_1 came to be able to produce the S_2-connected constellation of responses. Expectancy cannot be part of the mechanism of the performance of which it is a product.

Now if, in the Blodgett type of experiment, S_1 stands for choice-point stimuli and S_2 for goal-box stimuli, unrewarded runs will enable S_1 to evoke S_2-connected responses. With the first introduction of food into the goal box a new set of responses (characteristic of feeding) will be conditioned to S_2, so that at the next run a changed pattern of S_2-connected responses will be evoked by S_1. The expectancy-producing responses may be assumed to share in this change, but, as argued above, cannot be part of the causal chain determining the change in the *immediate responses* to S_1.

But may not S_2-connected responses be intermediaries in the greatly increased tendency for S_1 to evoke the appropriate turn? Seward (13, 14) has an ingenious theory that has potentialities for answering this question, and has promised to show how it "takes care of the phenomena of latent learning and of sudden changes in amount of rewards" (13, p. 372). A central idea of Seward's theory is that once a right turn at the choice point has been followed by reward in the goal box, at the next arrival at the choice point a kind of secondary drive ("tertiary motivation") produced by the proprioceptive cues from *starting* to turn right (VTE) preferentially facilitates right-turning (14). It seems precarious to lean so heavily on proprioceptive cues in view of the unimpaired maze performance of animals in which gross lesions have grossly altered proprioceptive impulse patterns (6, 7).

A NEUROPHYSIOLOGICAL VIEW OF LATENT LEARNING

It is rational to turn for an explanation of latent learning to the place where it occurs—the nervous system. Thirteen years of elaborate discussions about

[1] The way Tolman defines expectancy is not objective if objectivity in a definition implies that its referent is at least potentially observable by the observer and therefore part of the causal continuum from the observer's point of view. The cognitions of the subject can never be observed. Only the neural processes that are the correlates of these cognitions have the potentiality of filling the gap in the observer's causal continuum.

theory construction have left as firmly grounded as ever Pratt's assertion: "Only physiology can afford an explanatory frame of reference for the facts of psychology" (11, p. 131).

The hypothesis of latent learning to be outlined here is an extension of the neurophysiological hypothesis of learning previously advanced (22, 24). According to this hypothesis learning depends upon the development of functional connections (i.e., synapse formation) between neurons in anatomical apposition. On a basis of experimental evidence certain relationships were deduced which will be set down here in the form of postulates.

Postulate 1: The relevant presynaptic neuron must deliver impulses to the synaptic zone during, or just before, firing of the postsynaptic neuron. This provides sensitizing preconditions which we shall call *reinforcement-sensitivity.*

Postulate 2: Drive reduction must follow the formation of reinforcement-sensitivity if the process of synapse formation is to be completed. (Drive reduction means reduction in the number of impulses arriving at this synaptic zone from all sources. All stimuli are possible sources of impulses, and the number of impulses varies directly with strength of the stimuli. The reduction in number of impulses following the cessation of a weak visual stimulus would obviously be much less than would follow from marked relief of hunger.)

Postulate 3: Amount of synapse formation (and therefore amount of learning) varies directly with amount of drive reduction, other things being equal.

Postulate 4: The shorter the interval between the formation of reinforcement-sensitivity and the ensuing drive reduction, the greater the amount of synapse formation. Reinforcement-sensitivity fades with time.

To the above an additional postulate is now added:

Postulate 5: Amount of synapse formation varies directly with amount of reinforcement-sensitivity upon which the drive reduction acts.

Let it be assumed that in a latent-learning experiment of the Blodgett type a rat is eventually to receive food in the right compartment of a T maze. Thus, the habit to be built up will be a right-turning (R_r) in response to the cues of the choice point (S_{cp}). With reference to Fig. 1 we shall first discuss the hypothetical neural events associated with an unrewarded run in which the animal does a right turn at the choice point, and then the consequences of introducing reward.

FIG. 1.

Let n_1 represent a neuron leading to the response of turning right (R_r). During an unrewarded run, while n_1 is firing, impulses are being delivered to it by n_2 (from choice-point stimulation S_{cp}), by n_3 (from hunger stimulation S_h), and by n_4 (representing miscellaneous stimulations S_x). For n_2 to deliver impulses while n_1 is discharging them means that sensitivity for reinforcing the $n_2 \cdot n_1$ connection is created, and the amount of reinforcement that finally results will depend on (*a*) how much reduction of drive follows in the whole synaptic zone and (*b*) how soon it follows. During this unrewarded run the only drive reductions to occur are those correlated with the cessation of action of weak stimuli such as S_{cp} and S_x. Thus reinforcement is weak, and manifest learning is weak.

To account for latent learning the following supplementary hypothesis is suggested. A run followed by an adequate food reward would lead to sufficient drive reduction to use up *all* the

reinforcement-sensitivity produced at $n_2 \cdot n_1$ during that run. On the other hand, the weak drive reduction following an unrewarded run *leaves a residuum of reinforcement-sensitivity.* Part of this residuum fades, and *part remains in a cumulable form* that is activated by, and summates with, the new reinforcement-sensitivity produced during the next run. After a number of unrewarded runs there will be a considerable accumulation of reinforcement-sensitivity at $n_2 \cdot n_1$, and now a food-rewarded run will strengthen inordinately the $n_2 \cdot n_1$ functional connection (see postulate 5 above).

Though based on the typical Blodgett experiment the above hypothesis is also in complete harmony with the variations thereof. *Whenever* reward at last closely follows a previously unrewarded stimulus-response sequence, reinforcement will be strikingly great, and it does not matter if the response is a turning into a blind (4), if the reward is given outside the maze (9), or if the unrewarded runs are variably interrupted (20). In the last-mentioned case relatively less striking postreward learning would be expected at the more advanced choice points where the animal would have passed fewer times.

Corollaries

The following corollaries emerge from the above hypothesis:

1. *The greater the number of preliminary unrewarded runs the greater will be the increment in learning following the first rewarded runs.*

Blodgett's own experiment (2) seems to confirm this. The mean errors of his 3-day-unrewarded group decreased from 2.7 to 0.3 after four rewarded runs, while those of his 7-day-unrewarded group decreased from 2.3 to 0.2 after two rewarded runs. The mean improvement per rewarded run was thus 0.6 for the 3-day group and 1.15 for the 7-day group. The curve of improvement following the first reward flattens out for the 3-day group while remaining very steep for the 7-day group. This suggests that the accumulated reinforcement-sensitivity was entirely absorbed in the course of the first big drive reduction in the case of the former group but not in the case of the latter. Presumably, the amount of drive reduction to follow one reward of the size given was insufficient to use up the larger amount of reinforcement-sensitivity accumulated by the 7-day group. This presumption could be tested by repeating Blodgett's experiment with varying amounts of reward for different groups.

2. *The less the manifest learning that follows the unrewarded runs the greater the increment of learning that will follow the first few rewarded runs.*

The difference between the results of Blodgett (2) and those of Reynolds (12) seems to support this corollary.

Two Problems:

Does cumulable reinforcement-sensitivity (latent learning) slowly fade with time, and do reciprocal-inhibition effects (25) weaken it?

REFERENCES

1. BIRCH, H. G., & BITTERMAN, M. E. Sensory integration and cognitive theory. *Psychol. Rev.*, 1951, **58**, 355–361.
2. BLODGETT, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univer. Calif. Publ. Psychol.*, 1929, **4**, 113–134. In N. L. Munn, *Handbook of psychological research in the rat.* Boston: Houghton Mifflin, 1950.
3. BROGDEN, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, **25**, 323–332.
4. HERB, F. H. Latent learning—non-reward followed by food in blinds. *J. comp. Psychol.*, 1940, **29**, 247–256.
5. HILGARD, E. R. *Theories of learning.* New York: Appleton, 1948.

6. LASHLEY, K. S., & BALL, JOSEPHINE. Spinal conduction and kinaesthetic sensitivity in the maze habit. *J. comp. Psychol.*, 1929, **9**, 71–106.

7. LASHLEY, K. S., & MCCARTHY, D. A. The survival of the maze habit after cerebellar injuries. *J. comp. Psychol.*, 1926, **6**, 423–433.

8. MALTZMAN, I. The Blodgett and Haney types of latent learning experiment. *Psychol. Bull.*, 1952, **49**, 52–60.

9. MEEHL, P. E., & MACCORQUODALE, K. Personal communication to J. P. Seward. Experimental evidence for the motivating function of reward. *Psychol. Bull.*, 1951, **48**, 130–149.

10. MILLER, N. E. Comments on multiple-process conceptions of learning. *Psychol. Rev.*, 1951, **58**, 375–381.

11. PRATT, C. C. *The logic of modern psychology.* New York: Macmillan, 1939.

12. REYNOLDS, B. A repetition of the Blodgett experiment on "latent learning." *J. exp. Psychol.*, 1945, **35**, 504–516.

13. SEWARD, J. P. Secondary reinforcement as tertiary motivation: a revision of Hull's revision. *Psychol. Rev.*, 1950, **57**, 362–374.

14. SEWARD, J. P. Delayed reward learning. *Psychol. Rev.*, 1952, **59**, 200–201.

15. SPENCE, K. W. The differential response in animals to varying stimuli within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

16. THISTLETHWAITE, D. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, **48**, 97–129.

17. THISTLETHWAITE, D. Reply to Kendler and Maltzman. *Psychol. Bull.*, 1952, **49**, 61–71.

18. TOLMAN, E. C. *Purposive behavior in animals and men.* New York: Appleton, 1932.

19. TOLMAN, E. C., & HONZIK, H. C. Introduction and removal of reward, and maze performance of rats. *Univer. Calif. Publ. Psychol.*, 1930, **4**, 257–275.

20. WALLACE, S. R., JR., BLACKWELL, M. G., JR., & JENKINS, I. Pre-reward and post-reward performance in the "latent learning" of an elevated maze. *Psychol. Bull.*, 1941, **38**, 694. (Abstract)

21. WOLPE, J. An interpretation of the effects of combinations of stimuli (patterns) based on current neurophysiology. *Psychol. Rev.*, 1949, **56**, 277–283.

22. WOLPE, J. Need-reduction, drive-reduction, and reinforcement: a neurophysiological view. *Psychol. Rev.*, 1950, **57**, 19–26.

23. WOLPE, J. The formation of negative habits: a neurophysiological view. *Psychol. Rev.*, 1952, **59**, 290–299.

24. WOLPE, J. The neurophysiology of learning and delayed reward learning. *Psychol. Rev.*, 1952, **59**, 192–199.

25. WOLPE, J. Primary stimulus generalization: a neurophysiological view. *Psychol. Rev.*, 1952, **59**, 8–10.

[MS. received September 5, 1952]

Psychological Review
Vol. 60, No. 5, 1953

AN APPLICATION OF HELSON'S THEORY OF ADAPTATION-LEVEL TO THE PROBLEM OF TRANSPOSITION [1]

HENRY JAMES
University College, London

One of the usual requirements of a theory of behavior, psychological or physiological, is that it shall give a step-by-step account of the successive coordinations and differentiations, both sensory-sensory and sensory-motor, which lift behavior above the level of the schematic reflex. If this criterion is applied to the Gestalt and S–R theories of stimulus equivalence, both are found wanting. The neurophysiological theory of Köhler (17), preoccupied perhaps with shifting the problem of "similarity" from the periphery to the central nervous system, gives no account of the events which link the motor system with the cerebral activity aroused by the stimulus. The same objection may be raised against the recent use of such psychological constructs as "differentiation" (28) and "mediating process" (21) to account for certain instances of equivalence behavior; until something is known of how the organism may at once be influenced by such general states *and* respond to a particular stimulus in a particular way on any one trial, this kind of model raises more problems than it solves. Hull's (12) theory of stimulus equivalence, on the other hand, is, in a sense, more precise concerning the train of events from receptor to muscle (provided one casts a blind eye on "afferent neural interaction"). We may doubt the usefulness, however, of arguing back, as acceptance of the theory would force us to, from an observation of similar responses to the existence of sets of mediating stimulus continua along which generalization may take place. The existence of a continuum is inferred from certain established relations between measurable differences in stimuli presented and responses observed. Such complex stimuli as geometrical forms and nonsense syllables cannot be scaled independently of *S*'s response to them, and hence the psychophysical relations necessary for the study of similarity within the framework of Hull's theory cannot be determined.

If we accept Razran's (24) recent review of the literature, the study of stimulus equivalence at the level of Pavlovian irradiation throws us back upon the same problems. He concludes that the *S* in a classical generalization experiment forms some sort of crude rating scale of similarity against which the new stimuli are assessed. We are thus, it seems, forced to assume that a central process, of whose properties we know little, intervenes between the receptor stimulus and the response, which leads us back to the question of how these "sets" determine a particular response on a particular trial.

Similar problems have arisen from the traditional treatments of visual space perception; Gestalt theory again ignores the question of response, and functionalism provides no account of the way in which the various cues are integrated to produce a stable world. A way out of this dilemma is indicated by Gibson's recent suggestion that the problem of integration should be kept at the periphery. "If," he writes, "the total stimulation contains all that is needed to account for visual perception, the hypothesis of sensory integration is un-

[1] I am grateful to Messrs. A. Summerfield and A. Jonckheere for their helpful criticisms of this paper in manuscript.

necessary" (6, p. 25). While this type of analysis has certain limitations, its virtue is that it no longer requires us to map out the train of events which intervene between the receptor stimulus and the end response; rather it demands a psychophysical model in the form of sets of quantitative statements of relations between measurable properties of stimuli and some convenient metric of response frequency or extent. Ideally, the parameters of the mathematical model do not require, although they permit, translation into psychological or physiological constructs; but the limitations which the psychophysical approach places upon the range of appropriate stimulus situations, together with the absence of a purely rational basis for deriving the necessary equations, make this ideal scarcely realizable. Rather, as Graham points out, "the application of psychophysical method is a first step in setting up the conditions that allow for a scientific description of a stimulus-response relation. Properly conceived, it specifies variables. It also poses a relation between variables requiring explanation" (7, p. 117).

Psychophysics, in the sense in which Gibson uses the term, has not found much favor with students of learning, although Pratt's (23) paper on the "Law of Disuse" gave a clear indication of its usefulness in this field. In the following pages we shall attempt to show that by suitably adapting the method to the needs of the learning situation, it is possible to avoid the sort of problem outlined above, which arises when the more conventional hypothetical-construct–intervening-variable type of model forms the basis of our description of the mechanisms of stimulus equivalence.

ADAPTATION LEVEL AND TRANSPOSITION

The transposition experiment consists essentially in training S to respond differentially to two stimuli, X and X', the critical characteristic of which lies along a single physical continuum. The S is then presented with two new stimuli, one of which may be X or X', and rewarded for whichever choice he makes. The problem is to determine whether he has learned to respond to the absolute properties of the training stimuli or to the relation between them. The main facts which have been established in this sort of experiment are as follows:

1. Within a certain range of stimulus variation, transposition tends to occur (i.e., the test response is to the relative rather than to the absolute values).
2. Transposition tends to decrease as the distance of the test from the training stimuli increases.
3. Transposition tends to improve as testing is continued, even though S is rewarded for whichever response he makes.

These facts, suitably expressed in terms of the stimuli used and the frequency of the responses observed, constitute the main psychophysical relations for which any theory of transposition has to account. In terms of the present approach, the problem is to give them a compendious quantitative expression, and to this end we shall consider Helson's theory of adaptation level (9, 10, 11). The reader is referred to the original articles for a detailed exposition of the theory. In essence it assumes that:

> . . . effects of stimulation form a spatio-temporal configuration in which order prevails. For every excitation-response configuration there is assumed a stimulus which represents the pooled effect of all the stimuli [with which the S is or has been presented] and to which the organism may be said to be attuned or adapted. Stimuli near this value fail to elicit any response from the organism or bring forth such neutral responses as indifferent, neutral, doubtful, equal or the like, depending on the context of stimulation. Such stimuli are said to be at adaptation level. Conversely, adaptation level may be quantita-

tively specified by giving a value of stimulus eliciting the neutral response.

Stimuli above the adaptation level (AL) are assumed to establish positive gradients with responses of one kind, and stimuli below establish negative gradients with responses of an opposite kind. Positive or negative enhancement occurs as stimuli are above or below the AL; which may be high, medium or low, depending upon a number of factors in the stimulus configuration like background in vision, comparison stimulus in psychophysical judgments and in the organism which may be "set" at a given level due to effects of previous stimulation. A stimulus far above the stimulus range may raise the AL so high that most of the stimuli will be below the AL and the majority of judgments will be of the "negative" sort, while the opposite will be the case with a stimulus far below the range (10, p. 2).

In applying this hypothesis to transposition behavior we make the following additional assumptions:

1. In the process of training a response, say to the brighter of two stimuli, there is established in *S* a neutral point or region, corresponding to the AL of Helson.

2. The *S* learns to avoid stimuli which are less intense than the neutral point and to approach those which lie above it.

3. The process of establishing a neutral point is independent of reward and punishment. While the latter may determine the character of the responses which stimuli on either side of the neutral point will evoke, they are assumed not to contribute to the location of the neutral point itself.

4. New stimuli presented to *S* after training will induce a gradual change in the position of the neutral point. The rate of shift probably depends on such factors as number of trials with the training stimuli (13), distribution of practice, intelligence, etc.

From these assumptions it follows that if the test stimuli fall one on either side of the training AL, perfect transfer will be predicted. If both fall on the same side of the training AL, then responses will be random with respect to the critical cue until continued familiarity with the test stimuli has shifted the AL in the appropriate direction. In checking this point of view against some of the published literature (1, 3, 15, 25), we have assumed that the test stimuli have an immediate effect upon the location of the neutral point; that is, by analogy with the case of psychophysical judgment, we have assumed that the training neutral point has the function of a standard which initially "anchors" the transposition responses.

The formula found adequate by Helson for predicting the AL in weight-lifting and constancy experiments was a weighted geometric mean of the background (standard) and sample (series) stimuli. To test the hypothesis against rat data from Kendler (15) and Ehrenfreund (3) we have used the equation

$$\text{Log AL} = \tfrac{1}{2}\left(\frac{\Sigma \log X}{n} + \log C\right) \quad (1)$$

where AL is the neutral point for the initial test situation, X is the test stimuli, n their number, and C is the training neutral point (estimated as the geometric mean of the training stimuli). For Spence's (25) data on transposition in chimpanzees,

$$\text{Log AL} = \left(\frac{\Sigma \log X + \log C}{n+1}\right) \quad (2)$$

and for Alberts and Ehrenfreund's (1) experiment with preverbal children,

$$\text{Log AL} = \tfrac{1}{4}\left(\frac{3\Sigma \log X}{n} + \log C\right). \quad (3)$$

The relative weights given to the training and test stimuli in the above formulae are approximate but not altogether arbitrary, since it is perhaps reasonable to assume that as one moves up the evolutionary tree, the relative in-

TABLE 1

DATA FROM KENDLER (15). AL CALCULATED FROM EQUATION (1)

(In this and the succeeding tables, where prediction is at variance with the observation, the calculated AL is printed in boldface.)

Test Stimuli*	Per Cent Transposition			Calculated AL
	1st Trial	1st 5 Trials	2nd 5 Trials	
Condition I. $N = 58$. Training stimuli:* .063 (+), .011 (−)				
.37, .063	80%	80%	86%	.063
.37, 1.62	70%	72%	80%	.14
8.49, 1.62	60%	54%	80%	.31
40.46, 8.49	80%	72%	78%	.70
204.74, 41.22	67%	70%	92%	1.55
Condition II. $N = 24$. Training stimuli:* 8.49 (+), 40.46 (−)				
1.62, 8.49	67%	77%	82%	8.29
.37, 1.62	33%	60%	79%	3.79
.063, .37	67%	60%	72%	1.68
.011, .063	50%	63%	69%	.70

* Brightness in apparent foot-candles.

TABLE 2

DATA FROM EHRENFREUND (3). AL CALCULATED FROM EQUATION (1)

Test Stimuli*	Per Cent Transposition		Calculated AL
	Trials 1–5	Trials 16–20	
Experiment I. $N = 40$. Training stimuli:* .22 (+), .91 (−)			
.13, .57	86%	94%	.35
.09, .34	82%	74%	.28
.05, .22	48%	40%	.22
.035, .13	66%	54%	.17
Experiment II. $N = 40$. Training stimuli:* .13 (+), .035 (−)			
.05, .22	84%	80%	.08
.09, .34	78%	86%	.11
.13, .57	28%	28%	.136
.22, .91	40%	50%	.17

* Diffuse reflection factor.

fluences of the presented situation and of previous learning will change in favor of the former. The training and test stimuli used in the experiments mentioned above, together with the percentage of relational responses observed, and the calculated test AL are given in Tables 1–4. The analysis is further condensed in Table 5 by considering those cases in which the predicted AL falls between and those in which it falls outside the test stimuli against the occasions on which more or less than 80% of the observed responses on the first five trials were in the direction of training.[2]

DISCUSSION

While the agreement between observation and prediction in the tables does

[2] Spence does not give data for the first five trials in his experiments. His results, which are for ten trials, have nevertheless been included in Table 5.

TABLE 3

DATA FROM SPENCE (25). AL (INDICATED IN PARENTHESES) CALCULATED FROM EQUATION (2)

Training Stimuli* and Subjects	Test Stimuli,* Percentage of Responses, and Calculated AL†		
160 (+), 100 (−)	256, 160	320, 200	409, 256
Pati	100% (182)	80% (201)	80% (250)
	100% (180)	100% (209)	60% (255)
Mona	50% (173)	40% (208)	80% (246)
	100% (183)	70% (212)	60% (255)
Pan	80% (173)	90% (208)	90% (246)
	100% (183)	100% (212)	100% (255)
256 (+), 409 (−)	160, 256	100, 160	
Soda	50% (237)	40% (167)	
	70% (219)	50% (161)	
Bentia	40% (221)	30% (173)	
	60% (227)	40% (166)	
256 (+), 128 (−)	512, 256	128, 64	
Cuba	90% (287)	100% (120)	
Lia	100% (287)	100% (120)	
256 (+), 160 (−)	409, 256	160, 100	
Mimi	90% (277)	80% (153)	
128 (+), 256 (−)	256, 512	64, 128	
Jack	80% (273)	50% (114)	
Nira	100% (273)	60% (114)	
160 (+), 256 (−)	256, 409	100, 160	
Bokar	80% (267)	50% (148)	

* Area in square centimeters.

† The reader should consult Spence's article for details of the order in which the transposition tests were given. The AL's were calculated on the assumption that each test and each retraining series shifted the previous test or retraining AL to the same degree, irrespective of the number of trials given. It was also assumed that Cuba, Lia, and Mimi were tested first up and then down the size dimension, and that Jack, Nira, and Bokar were tested down and then up the scale.

TABLE 4

DATA FROM ALBERTS AND EHRENFREUND (1). AL CALCULATED FROM EQUATION (3) $N = 22$. Training stimuli:* 64 (+), 128 (−)

Test Stimuli*	Per Cent Transposition			Calculated AL
	1st Trial	1st 5 Trials	All 10 Trials	
64, 32	100%	100%	100%	53.8
32, 16	33%	73%	78%	31.8
8, 4	50%	53%	60%	11.3
4, 2	50%	50%	57%	6.7

* Area in square inches.

TABLE 5

SUMMARY OF TABLES 1–4. OBSERVED AGAINST PREDICTED FAILURE OR SUCCESS OF TRANSPOSITION IN FIRST FIVE TRIALS

No. of Tests in Which Per Cent of Transposition Responses Observed	Transposition Predicted	Failure of Transposition Predicted	Total
≥80%	23	4	27
<80%	14	18	32
Total	37	22	59

χ^2 (Yates's correction) = 9.1; $df = 1$; $p = <.01$.

not merit any extravagant conclusions, it is perhaps sufficiently close to warrant brief discussion. With the introduction of further parameters the correspondence could, no doubt, be improved in some cases, though in others (e.g., Mona and Pan, who, in spite of receiving the transposition tests in identical order, managed nevertheless to give divergent results [25]) this might not be possible without a greater knowledge of the *S*'s previous history. Since we are not interested in transposition as such, but rather as a special case of equivalence behavior to which a particular kind of analysis seems appropriate, we shall not pursue the matter of precision further.

Next we shall consider the assumptions we have made in transferring Helson's model from its perceptual context to that of learning. Most exception will probably be taken to the assumption that discrimination learning is a two-stage process. While this is not the place to enter into the current controversy on this issue (see, for example, 2, 14), we may perhaps make a point to which recent writers have not given much attention, and which is directly relevant to the hypothesis advanced here. There is ample evidence, in certain restricted fields of human activity at least, that some form of behavior modification takes place in the absence of either reward or knowledge of results. Perhaps the clearest examples of this are to be drawn from the changes which repetition brings about in subjective scales (13, 27) in visual illusions (18, 19, 22), and in the skills studied by Gagné and his associates (4, 5). Whether we call this learning or not does not seem to be of much consequence. Inasmuch as psychophysical experiments are essentially concerned with discrimination behavior, and as they have produced evidence that predictable changes occur in this behavior which cannot be ascribed either to drive reduction or to sign-learning, it becomes not unreasonable to suppose that such changes continue to take place when the discrimination situation also involves the learning of some new set of responses. The assumption is necessary in the present instance to account for the continued increase in the number of transposition responses in the test trials, where every response is rewarded (1, 15, 16).[3] The other assumptions are additional to those of Helson only in the sense that they translate his postulates into a learning terminology; as such they do not require further discussion.

Finally, something should be said of the possible extensions of the approach advocated here. We shall concern ourselves first with deductions from the hypothesis. This has already been tailored to fit three established sets of facts: the occurrence of transposition, its diminution with the distance of test from training stimuli, and the increase of relational responses with repeated testing. One deduction has already been mentioned: that if the test AL coincides with the training AL, perfect transfer should result.[4] A second is to the effect that in cases where transposition is initially less than perfect, but improves with testing, there will be an inverse relation between the number of test trials

[3] This fact is not readily accounted for by Spence's (25) theory. On the other hand, granted the assumption of generalization gradients, Spence (26) accounts nicely for the difficulty of the intermediate size problem, which we are unable to do at present.

[4] Since this paper was written, D. H. Lawrence (*J. comp. physiol. Psychol.*, 1952, 45, 511–516) has reported data which are in agreement with this inference. His test stimuli were brightnesses of 31.8 and 25.9 apparent foot-candles. The AL's of his ATG1, GTG, and ATG2 groups, calculated from equation (1), are estimated for these test stimuli at 31.92, 30.7, and 30.37, respectively. Positive transfer is therefore predicted for the last two groups and negative transfer for the first.

carried out and the number of "correct" responses observed when one switches back to the original training pair. More complex effects, corresponding to the establishment of learning sets (8; 20, p. 129), will presumably result if this process is carried out too often.

Before the theory can be extended beyond such elementary, first-order deductions, however, it will probably be necessary to fill certain gaps in our knowledge. One of the more obvious is to define the limits of improvement which will occur with testing. A pinhead and a cartwheel do not lie on the same psychological dimension, and an examination of the relation between these discontinuities and the value of the anchoring point would not be without interest. Another problem is to determine, under a number of conditions, the relation between the amount of practice given on the training stimuli and the rate at which the AL shifts. This is of some importance if the present theory is to be extended to cover the intermediate size problem, which takes animals much longer to learn than does a conventional discrimination (26, 29).

At a more general level, it should be pointed out that the AL concept makes nonsense of any attempt to examine the basis of equivalence by the successive presentation of different stimuli under conditions where the sequential effect cannot be taken into account. If every stimulus alters the frame of reference against which the next is judged, it would seem that the difficulties confronting a qualitative analysis of this process are at present insuperable.

SUMMARY

1. It is argued that a psychophysical theory of stimulus equivalence avoids certain problems of sensory-sensory and sensory-motor integration which arise out of the application of S–R and Gestalt theories to this area.

2. It is shown that the main facts of transposition behavior, as a simple case of stimulus equivalence, can be derived from an extension of Helson's theory of adaptation level.

3. Some further problems suggested by this approach are briefly discussed.

REFERENCES

1. ALBERTS, E., & EHRENFREUND, D. Transposition in children as a function of age. *J. exp. Psychol.*, 1951, **41**, 30–38.
2. BIRCH, H. G., & BITTERMAN, M. E. Reinforcement and learning: the process of sensory integration. *Psychol. Rev.*, 1949, **56**, 292–308.
3. EHRENFREUND, D. A study of the transposition gradient. *J. exp. Psychol.*, 1952, **43**, 81–87.
4. GAGNÉ, R. M., & FOSTER, HARRIET. Transfer to a motor skill from practice on a pictured representation. *J. exp. Psychol.*, 1949, **39**, 342–354.
5. GAGNÉ, R. M., & BAKER, KATHERINE E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
6. GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.
7. GRAHAM, C. H. Behavior, perception, and the psychophysical methods. *Psychol. Rev.*, 1950, **57**, 108–120.
8. HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
9. HELSON, H. Fundamental problems in color vision. I. The principle governing changes in hue, saturation, and lightness of non-selective samples in chromatic illumination. *J. exp. Psychol.*, 1938, **23**, 439–476.
10. HELSON, H. Adaptation level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
11. HELSON, H. Adaptation level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.
12. HULL, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46**, 9–30.
13. JOHNSON, D. M. Learning function for a change in the scale of judgment. *J. exp. Psychol.*, 1949, **39**, 851–860.
14. KENDLER, H. H., & UNDERWOOD, B. J. The role of reward in conditioning theory. *Psychol. Rev.*, 1948, **55**, 209–215.

15. KENDLER, TRACY S. An experimental investigation of transposition as a function of the difference between training and test stimuli. *J. exp. Psychol.*, 1950, **40**, 552–562.
16. KLÜVER, H. *Behavior mechanisms in monkeys.* Chicago: Univer. of Chicago Press, 1933.
17. KÖHLER, W. *Dynamics in psychology.* New York: Liveright, 1940.
18. KÖHLER, W., & FISHBACK, JULIA. The destruction of the Müller-Lyer illusion in repeated trials: I. An examination of two theories. *J. exp. Psychol.*, 1950, **40**, 267–281.
19. KÖHLER, W., & FISHBACK, JULIA. The destruction of the Müller-Lyer illusion in repeated trials: II. Satiation patterns and memory traces. *J. exp. Psychol.*, 1950, **40**, 398–410.
20. LASHLEY, K. S. The mechanism of vision: XV. Preliminary studies of the rat's capacity for detail vision. *J. gen. Psychol.*, 1938, **18**, 123–193.
21. LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.
22. LEWIS, E. O. The effect of practice on the Müller-Lyer illusion. *Brit. J. Psychol.*, 1908, **2**, 294–306.
23. PRATT, C. C. The law of disuse. *Psychol. Rev.*, 1936, **43**, 83–93.
24. RAZRAN, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.
25. SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
26. SPENCE, K. W. The basis of solution by chimpanzees of the intermediate size problem. *J. exp. Psychol.*, 1942, **31**, 257–271.
27. TRESSELT, M. E. The influence of amount of practice upon the formation of a scale of judgment. *J. exp. Psychol.*, 1947, **37**, 251–260.
28. UNDERWOOD, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**, No. 3 (Whole No. 273).
29. WOLFLE, D. L. Absolute brightness discrimination in the white rat. *J. comp. Psychol.*, 1937, **24**, 59–71.

[MS. received October 2, 1952]

Vol. 60, No. 6 November, 1953

THE PSYCHOLOGICAL REVIEW

THE SCIENCE OF PERSONALITY: NOMOTHETIC OR IDIOGRAPHIC? [1]

SAMUEL J. BECK

University of Chicago, Northwestern University, and Michael Reese Hospital

A principal and vigorously debated issue before psychology today concerns method, and the adequacy of the known methods of evaluating findings by any of the approaches to the study of the whole personality. The ferment centers especially around the merits of the "nomothetic" as against the "idiographic" logic. First, let it be noted that, so far as concerns the basic procedures of scientific method, the two methods have everything in common. They both have recourse to observation and to experiment. They analyze and resynthesize data. They draw inferences that follow the usual canons of logic, both inductive and deductive. These are the foundational approaches to scientific method. The two procedures—the nomothetic and the idiographic—do differ, however, and critically. The differences have weighty effect in the kinds of hypotheses that can be formulated concerning human behavior. Such effects can therefore be very important for progress in the study of the human personality, since hypotheses are really the product of scientific imagination. The issue becomes: how liberated is the scientific imagination, always the starting point for new experiment into unknown areas of any particular field?

Historically, this nomothetic-idiographic issue could hardly have become so important in psychology before the present time. The nomothetic method has so far been the traditional one. But interest in personality, as personality, is shaking the logic on which it rests. The logic is not proving adequate for psychologists who are focusing on this behavior unit at the same time that they want to adhere to a good method. They have now been hard put to it to find a logic applicable to the phenomenon which they have elected to study.

Some light on the inadequacy of the nomothetic procedure in the field of personality is obtainable by considering the questions that we want answered in such investigations. We may inquire how much of some one personality trait is found in each member of a population under study—bravery, for example, or pride, or sense of humor. Or we may ask about any person how much bravery does he have, *and* coolness, *and* pride, *and* sense of humor, *and* other variables that fuse into character. In the one, we ask what are the incidence and the

[1] Presidential address of the Division of Clinical and Abnormal Psychology, read at the Amer. Psychol. Ass., Washington, D. C., September, 1952.

The thinking in the present paper grew in part out of two researches in schizophrenia (MH 63 and MH 64) conducted under a National Institute of Mental Health grant, in the Psychosomatic and Psychiatric Institute, Michael Reese Hospital, Chicago, Illinois.

distribution of some one datum within a population. In the other, we focus on the various behaviors within one person, their mutual interaction, and their effects in bringing about the total behavior which we identify as a particular personality.

The first of these two ways has so far been the proven one of science. It came to us following long experience in measurement, successful prediction, with resulting control. It has been so successful that it has captured the interest of all scientists, and that is equivalent to saying that it has harnessed the scientists themselves. To harness is to bind. It is a circumstance which can have disadvantages, restrictions. This is the situation which dominated in the natural sciences for a time, then in the social sciences, and finally in psychology. To have captured means to have made captive. This can trammel the thinking, channelize research, and, by preventing new imagination, block progress.

Now psychology, as a science, is comparatively young. The science of the whole personality is the youngest child of this young science. It is characteristic of the young to look with reverence on, and worshipfully obey, their elders. So psychology and the study of personality obediently followed in the footsteps of their predecessors and attempted to build structures on the old pattern—nomothetic. Fortunately, children not only obey; they also rebel. However, in the present case I detect no hostility to the parental science figure, but a departure—a branching off. It is a sign of good health, in both parent and offspring, when the latter departs from familiar pathways, and the former can see him do it with equanimity.

SUB- AND SUPRAPERSONALITY RESEARCH

This datum, the whole personality, is, to a considerable group of investigators, including some in closely associated fields, the basic one in human behavior. It is the only *raison d'être* for the science of psychology. All other branches within this field are so many roads that lead to it. They do so from two directions.

One is at a "subpersonality" level; the other, "suprapersonality." In either event, they are significant sciences only in so far as they throw light on the behaving human. Among the sciences that lead to it from subpersonality levels are neurology, physiology, experimental psychology, psychometric tests. Specific researches that would be grouped under these heads would include those in conduction, the synapses, EEG, reflexes (unconditioned), all the receptor processes, psychophysics, the conditioned response, the functions and effects of such structures as the thalamus, adrenal cortex, autonomic nervous system, localization of function in the brain, various effects of nutrition, problems relative to sleep, etc. Some perception experiments belong here, although the trend now is to see perception as activity of the whole person.

Among those sciences that lead to the study of personality from the *supra* level are: anthropology, sociology in the older sense, social work as applied in community institutions, and the current newer disciplines known as group dynamics or action research. Here I would specify all studies in cultures as bearing on the personality or on formation of character; the developmental stages of the personality, including infancy, childhood and adolescence, maturity and old age, the special methods of studying the child—e.g., play techniques—would belong here. All the "functional" clinical pictures as, in fact, response to the environment would group here: the neuroses and the psychoses, and also the psychoanalytic formulations concerning them. Here

too belong the studies of special traits or behavior clusters, e.g., jealousy, suggestibility, masculinity, femininity, and sexual behavior. Also, here are such complex activities as delinquency, moral attitudes, aesthetics, religious systems. Finally, there is the individual in his industrial and business relations, in his political behavior, in the military services.

These lists do not pretend to be a systematic or exhaustive synopsis. They are more of the nature of random jottings. A complete statement would embrace all the efforts of psychologic science. Many psychologists will disagree. They will question whether some of these efforts properly belong within the curricula of a university department of psychology. Some are not quantifiable, or very slightly so; hence, are they science? The other horn of the psychologist's dilemma is: he collects data—rigorously controlled—enabling him to predict, but only within the narrow confines of the conditions set up in the experiment. How applicable are these to the essential problems of human life? It all comes down to the question as to what is the field proper of the science of psychology. It is a question on which there are both ponderings and searchings in important centers for the study of psychology.

To the student of the personality as a whole, the deficit in the nomothetic technique is that it dictates turning one's back on this unit datum. Variables are observed one by one. Each individual is measured in respect to only one such trait at a time, never in the interaction of the traits within him. The individual is atomized; and we stop with studying the atom. This is a product of the intellectualist temper, which has been under severe critical attack, especially from within philosophy (Lotze, James, Bergson). Psychology has been a closely guarded prisoner of the intellectualist temper.

Regarding the individual variable, some indirect light on its insufficiency comes from a recent paper on semantics. Rapoport (4), in the *American Scientist*, speaks of the "propositional function" of a statement. For example, the statement "the man is brave" looks like a logical proposition. But it cannot in itself be considered true or false unless the meaning of the predicate "brave," i.e., its function, is known. The man may be brave in battle . . .

> ". . . Jealous in honour, sudden and quick in quarrel,
> Seeking the bubble reputation
> Even in the cannon's mouth. . . ."
>
> . . . yet a coward in love (Miles Standish), . . .
>
> "I'm not afraid of bullets, nor shot from the mouth of a cannon,
> But a thundering 'NO,' point-blank, from the mouth of a woman,
> That, I confess, I'm afraid of."

He may eschew primitive hand-to-hand combat and be able to face the more deadly enmity of society (Mahatma Ghandi and many of the martyrs of religion).

Similar reasoning holds for any test that attempts to explore a whole personality, any multivariate instrument such as the Rorschach test. Any statement concerning any Rorschach test variable needs to be examined for its "propositional function." What we find out about a person in any one score is not necessarily true, and not necessarily false. We need to know its function in so far as that depends on its setting. We must therefore know that setting, the larger whole, if his behavior is to be predictable. At the present, the techniques used are Rorschach's test, Murray's TAT, other projective tools, the MMPI. Possibly some new kind of test, yet to be devised, will be more suc-

cessful. What ultimately happens to any now in use is not the important thing. It is important to have instruments that will dependably penetrate the whole person, as a psychodynamic system of stresses. Psychology can be depended on to devise them. It did so in psychometric science. The growth and now weighty influence of the Freudian hypotheses have given rise to the projective tests. But while the psychometric techniques did not present any challenging statistical difficulties, the projective instruments forthwith generated new problems—those of testing the test. You are familiar with the controversies. The next phase, in which we are now, was to throw the issue into the laps of the logicians of science.

THE USES OF IDIOGRAPHY

The new approaches are designated by the term "idiographic." The essence of these is that they permit the science of personality to keep in its syllogisms both terms, *science* and *personality*. A criticism—even stricture—of the nomothetic is that it can stick to science only at the expense of personality. In fact, the more scientific it is, the more isolated the variable it studies, and the less it takes in of the personality. But it is the human personality which is the objective of psychology. All the other sciences, to repeat, are *sub* or *supra*.

The essence of the idiographic method is that it focuses its glass on the universe of behavior traits co-functioning as a universe. This is the individual. But before it can be employed, a new question is legitimately posed and thrown back at the idiographist: how can we know—in the Rorschach test, for example—whether a productivity of, say, 30 or 50 is high or low? The same for an $F+$ per cent of so-or-so much, and the others. You cannot make the evaluations until we have criteria obtained by measuring each variable as distributed within a population. This is to fall back on the nomothetic approach. That is a first step. It must precede the idiographic.

The nomothetic is essential in setting up certain limits of quantitative findings. Doing this in psychology provides only certain directions of measurement, not absolute foot-rule figures. The directions of these measures in some populations—say normals or neurotics, or in children of one rather than another age level—and the differential directions in the other populations result in leads as to the organization of the variables in each clinical population. This organization or clustering is the raw material for a hypothesis to the effect that any one person within a clinical population would yield, in the variables sampled by the particular test, measures approximating certain limits. Conversely, a person found to measure in a number of these variables at or near certain quantitative limits belongs in the corresponding clinical population. At this stage, and for the purpose of experiment, the description of any person in the language of the psychological instrument used is itself a hypothesis. The next job becomes that of testing the hypothesis. This, much simplified, is the idiographic method. To test the hypothesis, we use all the usual methods of statistics. One of these, factor analysis, so employed becomes more directional than are its present habits. It circumscribes its individual tasks, does not take broadside shots, but "draws a bead," if you will, on certain objectives already found to be of potential significance.

To recapitulate, the steps in a scientific investigation of personality—after the test instrument has been chosen—are three. (*a*) The various behavior components accessible to the instrument are measured, and their distributions

plotted in the several clinical populations. This is nomothetic. From this step certain hunches emerge—tentatively we call them assumptions—as to the psychologic significance of the several components. (*b*) Individuals are studied by means of the test. Certain inferences as to dynamic relations between psychologic components within the individual follow—conclusions as to the mutual influences of the variables on one another and in producing the total picture. For example, lively imagination with too little appreciation of reality emerges as schizophrenia. The hypotheses that form immediately point the way to experiment, for the effect of the independent phenomenon on the dependent one: of schizophrenia, for example, on the pattern in a test such as the Rorschach. The third step is, therefore, (*c*) the experimental investigation itself. It obtains response patterns—the dependent structures—from individuals illustrating various independent conditions (neuroses, normal behavior, and the others) and tests these by the tried methods of investigation, correlation, factor analysis, analysis of variance, and the others. It is at point (*b*) that the method becomes idiographic.

THE LIMITATIONS ON NOMOTHESIS

The deficit of the nomothetic approach is that the more thorough and exact it is in adhering to its principles, the less adequate it is for exploring the whole functioning human unit. For, as it uses an ever larger and larger number of observations of one variable in different individuals, it reduces to a minimum the effects of other traits in the individual on that particular variable. So it obtains a datum that can be described independently of its shaping or distortion by each whole person. The greater the number of individuals in whom the trait was observed, the smaller is the standard deviation. So, the greater the "objectivity." We have eliminated the extravariable influences on the variable—extravariable, but intrapersonal—which is to say, we have eliminated the person. That is exactly what we want to do in such an experiment. The datum that is external to the rest of the personality is the datum of scientific interest of the moment. We take the person out of the observed datum, and concentrate only on the extraperson event. This is what scientific psychology is succeeding in doing. It is taking the human being out of the investigation. Any psychologist has the right to do so. Any investigator has the liberty of staking out his ground, defining his area, and limiting his objective. He must only be explicit in stating his definitions and his objective. What questions is he undertaking to answer?

The corollary that the unity of the person is a function of the brain acting as an organ, and an integrating one, is implicit in the thinking of some of the profound investigators of this organ. I mention here Sherrington, Herrick, and Lashley. A very recent paper by Sperry is apropos. His subject is "The Mind-Brain Problem." He concludes: "We should not expect to find that a single neuron or an isolated patch of neurons, or even a cortical center, could sense, feel, experience, or think anything in isolation. These psychic properties we envisage as depending upon a specific design and complexity in the vortex of neural activity, generally involving a reciprocal interplay of many parts" (5, pp. 310–311).

A universe of traits, variables in mutual interplay, affecting one another, these are the individual. This is the task which the idiographic method undertakes. The specific technique de-

vised to test out the findings in this kind of universe is that associated with Stephenson—the *Q* technique. It is to him, too, that I am indebted for the concept of the individual as a universe of traits. To quote his terse statement distinguishing the two methods, "Nomothesis was long regarded—and in effect, is—'psychology without a subject,' whereas idiography tried to concern itself with a self" (6).

The *Q* technique was the one I used in two researches in schizophrenia, one in adults and one in children (1947–1952). We have succeeded in these researches in isolating six schizophrenic reaction patterns. That is, we are describing six patterns within this disease group that differ from one another. They differ in the language of the Rorschach test items. They differ also in the corresponding clinical language. They make clinical sense. Full report of these researches will be published separately. The psychologist will be disappointed, however, who looks for descriptions of personality in terms of pure mathematics, logical deductions from quantitative findings, and from inferred relationships. Our findings do not provide such. Nor do I think that the science of personality by any technique has made progress such that it can so describe the personality. The science is in just too undeveloped and unsophisticated a state. I here quote another writer, Webster, discussing not the relatively complex field of personality as a whole, but psychology generally. He says: "Mathematical theory which is appropriate for interrelating many quantitative, continuous variables, is readily available, but there are always difficulties in applying it with any rigor to psychological data. This seems to be due mainly to the problems of measurement and possibly also to the fact that psychological variables may contain discontinuities which are poorly understood, and which may therefore vitiate the ordinary methods of analysis" (7, p. 168). What would be the ideal of such a science is stated by Boring in an article not *ad hoc,* but containing relevant comments: "The body of exact quantitative knowledge that we call science nowadays . . . the scientific psychology of the visual world, differs from phenomenology in being a collection of observed functional relations. . . . You cannot see the visual world at any moment when you are playing scientist; you construct it out of elaborate observations that have been collected for many years in the past" (1, p. 146).

The "discontinuities" of which Webster speaks are only too well known among the variables in which the clinical psychologist must deal. Hallucinations offer a case in point. Campbell, in one of his shorter papers (2), describing a girl of nineteen who appeared to be hallucinating, notes the difficulty of differentiating between her experiences as actual perceptions, or work of the imagination. "The patient does not make a consistent integration of different attitudes, but allows them to exist side by side," he says. Again: "Such a condition is not fixed and static; there is an oscillation between the different levels of thought. . . ." Such being the phenomenon which the clinical psychologist attempts to study, he does not have the aspiration as yet to describe them mathematically so as to answer Boring's criterion, "a body of exact, quantitative knowledge that we call science." In our idiographic study of schizophrenia, in patterning out the six types, we do not even claim that there is exact correlation between Rorschach test operations and the presumably corresponding clinical behavior. What we do say—and this is the meaning of our idiographic experiment—is that we can describe by means of the test the kind of whole person who, under certain

stress, or in one of Lewin's "field" conditions, is likely to have hallucinations.

Mention of Kurt Lewin's name recalls a quotation from him. Students in the field of personality, conscious as they are of the crudity of their instruments, of the long lag yet ahead before they can logically deduce constructs from mathematical findings, will find comfort in these words, coming from a man whose life was dedicated to the one goal, a science of personality. Lewin sees ". . . the basic character of science as the eternal attempt to go beyond what is regarded scientifically accessible at any specific time. To proceed beyond the limitations of a given level of knowledge, the researcher, as a rule, has to break down methodological taboos which condemn as 'unscientific' or 'illogical' the very methods or concepts which later on prove to be basic for the next major progress" (3, p. xv).

We may therefore expect more and more recourse to the idiographic approach, not only in psychology, but also in the other sciences concerned with human behavior. Idiography cannot do it alone. It must have a preceding groundwork by nomothesis. It is from the integrated use of the two methods that we will make progress. Out of such a synthesis should come a science of personality that will effectively retain both terms of this proposition: personality and science.

REFERENCES

1. BORING, E. G. Visual perception as invariance. *Psychol. Rev.*, 1952, **59**, 141–148.
2. CAMPBELL, G. M. Hallucinations: their nature and significance. *Amer. J. Psychiat.*, 1930, **9**, 607–618.
3. LEWIN, K. Quoted by D. Cartwright in *Field theory in social science.* New York: Harper, 1951.
4. RAPOPORT, A. What is semantics? *Amer. Scientist*, 1952, **40**, 123–135.
5. SPERRY, R. W. Neurology and the mind-brain problem. *Amer. Scientist*, 1952, **40**, 291–312.
6. STEPHENSON, W. *Introduction to Q-technique.* Chicago: Univer. of Chicago Press, 1952. (Duplicopied)
7. WEBSTER, H. Dynamic hypotheses in psychology. *Psychol. Rev.*, 1952, **59**, 168–171.

[MS. received October 2, 1952]

Psychological Review
Vol. 60, No. 6, 1953

LEARNING AND THE PRINCIPLE OF INVERSE PROBABILITY[1]

DAVID BAKAN
University of Missouri

The basic idea which has guided the theoretical explorations contained in this paper is that *science is a way of learning*. This thesis and some of its implications have been discussed in an earlier paper (1). In this paper the attempt is made to gain an understanding of the nature of the learning process through the examination of one formulation of the nature of the scientific method, the principle of inverse probability. This principle is best stated in its mathematical form, which will be found below. In brief, it formulates the effect on the probability of a theory of a confirmation. The principle formulates one important aspect of the scientific method. As such, it should tell us something concerning the psychology of learning.

In a sense, what follows can be considered to be the elaboration of Hull's "hidden" theory of learning; for Hull has not one but two theories of learning. One of these is that which is generally understood to be Hull's theory; the other is contained in his discussions on method, particularly the introductory sections of (8) and (9), and can be called the "hypothetico-deductive theory of learning." The latter is a theory of how the scientist learns. In this latter theory the three major concepts are *theory*, *observation*, and *probability*. Probability is what characterizes the relation between theory and observation. Thus, on the grounds of certain observations, a theory is said to have such and such a probability. On the grounds of a given theory, a predicted observation has such and such a probability of occurring. When a predicted observation is verified, the probability of the theory goes up. When a predicted observation fails to be verified, the probability of the theory goes down.

This "hidden" theory of learning is more of a cognitive than an S-R theory. It points almost directly to the principle of inverse probability, with which we shall deal below. It is to this principle that we turn for a fuller understanding of the hypothetico-deductive theory of learning and for its formalization.

The Principle of Inverse Probability in General

According to Jeffries (11), the principle of inverse probability "is to the theory of probability what Pythagoras's theorem is to geometry." Although the principle is one of the most critical bones of contention among probability theorists, it is mathematically sound. This is attested to by Uspensky, who says "Bayes' formula, and other conclusions derived from it, are necessary consequences of fundamental concepts and theorems of the theory of probability. Once we admit these fundamentals, we must admit Bayes' formula and all that follows from it" (17, p. 69). The major question is its applicability.

Probability theorists fall into two general categories, frequency theorists

[1] The writer is deeply indebted to the Research Council of the University of Missouri for a research professorship for the summer of 1952 which made the investigations reported in this paper possible.

and nonfrequency theorists.[2] The frequency theorist insists on an "objective" definition of probability in terms of relative frequency. The nonfrequency theorist considers probability to be a kind of rating scale of credence. Probability is, according to the nonfrequency theorist, "The state of mind with respect to an assertion, a coming event, or any other matter on which absolute knowledge does not exist" (4). The basic syntactical unit for the nonfrequency theorist is $P(g/h)$,[3] the probability of g on the grounds of h. If g necessarily follows from h, as in deduction, then $P(g/h) = 1$. If g is impossible on the data h, then $P(g/h) = 0$.

It is important to specify the point of view of this paper with respect to this controversy. On the philosophical level no position is taken. The acceptance of the principle is only on the grounds of its cogency with respect to what it might mean for the psychology of learning.

Mathematical Formulation of the Principle of Inverse Probability

The principle of inverse probability formulates the effect on the probability of a proposition g of the verification of a proposition x. Let us say that, on the basis of a previously acquired set of propositions h, a proposition g has the probability $P(g/h)$.

Then, say, a proposition x is found to be true. What is the effect of finding x to be true on the probability of g? A root form[4] of the principle (11) provides us with the answer:

$$P(g/hx) = \frac{P(x/gh)}{P(x/h)} P(g/h). \qquad [1]$$

The probability that g is true on the grounds of h and x is equal to the probability that g is true on the grounds of h, multiplied by the ratio of the probability of x on g and h, to the probability of x on h (h alone).

Thus, if x stems equally well from g and h, as from h alone, the probability of g does not change. If, however, x follows from g and h with a probability greater than the probability of x on h alone, the probability of g increases. Similarly, if $P(x/gh)$ should be less than $P(x/h)$, the probability of $P(g/hx)$ is less than $P(g/h)$.

The nature of this formula will become evident if we make the following interpretations of g, h, and x: Let h be a set of propositions concerning data which have been collected; let g be a theory which has been developed to account for the data h; and let x be a proposition about a new datum which was not involved in the initial generation of g. The formula then expresses the most critical aspect of the hypothetico-deductive method,

[2] Nagel (15) has recently summarized the controversy. Nagel himself accepts the frequency point of view and presents a rather weak case for the nonfrequency point of view. A balanced picture of the situation can be had by reading Nagel on the controversy in general, supplemented with, perhaps, Keynes's (13) presentation of the nonfrequency point of view.

[3] The symbol "/" should not be confused with this symbol as used to indicate division. It does not indicate division as used here. It should be read as "on" or "on the grounds of."

[4] This is immediately derivable from the theorem of compound probability. This theorem is (17):

$$P(ABC) = P(A)\,P(B/A)\,P(C/AB). \qquad (a)$$

We then have

$$P(hgx) = P(h)\,P(g/h)\,P(x/gh) \qquad (b)$$

and

$$P(hgx) = P(hxg) = P(h)\,P(x/h)\,P(g/hx) \qquad (c)$$

and

$$P(g/hx)\,P(x/h) = P(x/gh)\,P(g/h) \qquad (d)$$

and therefore

$$P(g/hx) = \frac{P(x/gh)}{P(x/h)} P(g/h). \qquad (e)$$

the effect of the confirmation of a deduction on the probability of the hypothesis that generated it. If x is found to be true, and x stems from the theory with a relatively high probability, then the probability of the theory is raised.

By appropriate algebraic manipulation,[5] formula [1] leads to

$$P_n = \frac{R\,P_{n-1}}{R\,P_{n-1} + (1 - P_{n-1})}, \quad [2]$$

where P_n is the probability of g after the verification of x_n, and P_{n-1} is the probability of g prior to the verification of x_n, and R is the ratio of $P(x_n/gh)$ to $P(x_n/\bar{g}h)$. ($\bar{g}$ = not $- g$.)

THE INTERPRETATION OF g, h, AND x

An interpretation of g, h, and x as propositions concerning theory, old data, and new datum, respectively, has already been given. However, for the principle of inverse probability to have greater generality for the psychology of learning, we will widen the interpretation of these symbols. The crux of this wider interpretation is in the interpretation of x, with which we shall deal first.

Let x stand for the conditional proposition, "If Y, then X," where X is an observation on the part of the learning organism, and Y the conditions for the observation. By the verification of x, we mean the verification of this whole conditional proposition. Some examples may be helpful:

Instrumental conditioning: x = "If I press the bar, I will get a pellet of food."

Classical conditioning: x = "If the metronome ticks, meat powder will be injected into my mouth."

Problem solving: x = "If I do such and such, the obstacle will be overcome."

Prejudice: x = "If he is a Negro, then I will find him to be irresponsible."

Law: x = "If this body is immersed in this liquid, then it will be buoyed up by a force which is equal to the weight of the liquid it displaces."

With this as our understanding of the meaning of x, *g is, then, whatever it is that generates x's.* g is that which is associated with the organism when we say that the organism has learned. It is the "what is learned" in the sense that we say that habits, attitudes, prejudices, skills, cognitions, hypotheses, etc. are learned. It should be emphasized that g is in no sense the overt behavior, but is, rather, the condition of the organism. The essential characteristic of g is that it can generate x's.

The h is whatever led to the g. It consists of hereditary factors, maturational factors, and previous learnings.

THE AMOUNT OF LEARNING AS A FUNCTION OF ONE POSITIVE "REINFORCEMENT"

"Reinforcement" may now be defined as the verification of x. We are now in a position to determine the amount of increase in the probability of g as a function of one reinforcement.

The amount of increase is given by the increment in the probability of g.

$$P_n - P_{n-1} = \frac{R\,P_{n-1}}{R\,P_{n-1} + (1 - P_{n-1})} - P_{n-1}. \quad [3]$$

[5] From formula [1] we have

$$P(g/hx) = \frac{P(x/gh)}{P(x/h)} P(g/h) \qquad (a)$$

and therefore also

$$P(\bar{g}/hx) = \frac{P(x/\bar{g}h)}{P(x/h)} P(\bar{g}/h). \qquad (b)$$

By dividing (a) by (b), with

$$P(g/hx) + P(\bar{g}/hx) = 1,$$

with $P(g/h) + P(\bar{g}/h) = 1$, and letting

$$\frac{P(x/gh)}{P(x/\bar{g}h)} = R,$$

we get

$$P(g/hx) = \frac{R\,P(g/h)}{R\,P(g/h) + [1 - P(g/h)]}. \qquad (c)$$

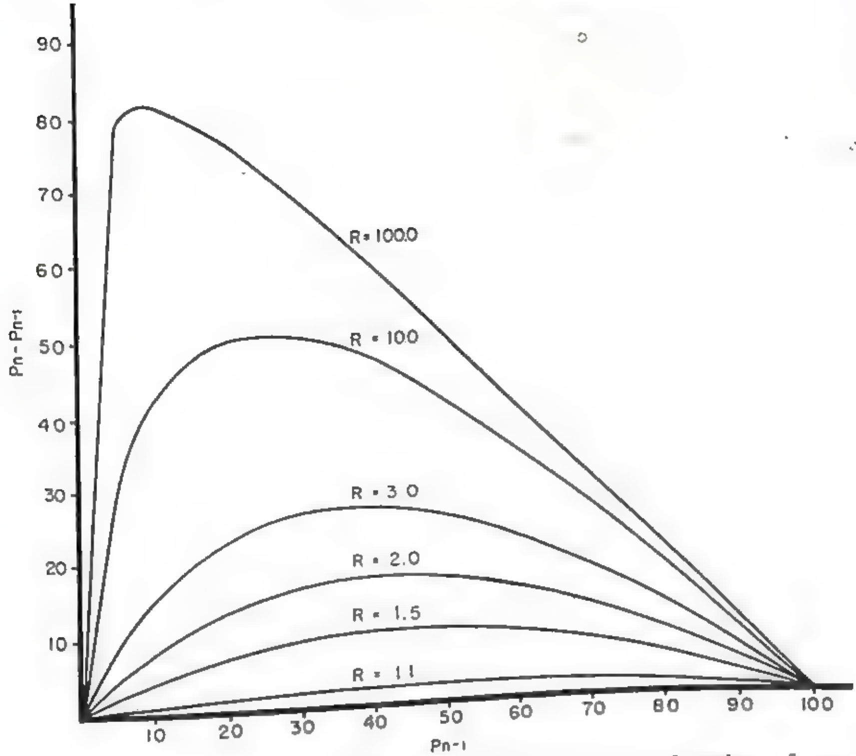


FIG. 1. The amount of increment, $P_n - P_{n-1}$, as a function of one "reinforcement," as a function of P_{n-1} and R

In Fig. 1, the values of the increment have been plotted for various values of R as a function of P_{n-1}. For any value of R which is greater than 1, the increment is low for both low and high values of P_{n-1}, and high for middle values of P_{n-1}.

Immediate empirical confirmation of the relationship indicated by this figure is to be found in phenomena of transfer of training. Roughly, where an individual approaches a "new" learning situation without already having very much to transfer to it, the initial learning of the new situation is slow, i.e., when P_{n-1} is low, $P_n - P_{n-1}$ is low. If, however, P_{n-1} is a middle value, improvement is rapid; and if P_{n-1} is already high, there is little more room for improvement. The work of Harlow (5) with respect to the learning of learning sets is a case in point. Translating his results into the terminology advanced here, where the initial probability of g is low, the increments as a function of reinforcement are low. Where the initial probabilities are of middle values, the increments are highest. Where the initial probabilities are high, the increments go down again. Or, what is the same thing, when P_{n-1} is low, learning is positively accelerated; and when P_{n-1} is high, learning is negatively accelerated.

The phenomenon of positive transfer of training can be further specified. From the mathematics of probability we have the formula: [6]

$$P(g_2) = P(g_1)\frac{P(g_2/g_1)}{P(g_1/g_2)}. \qquad [4]$$

[6]
$$P(g_1 g_2) = P(g_1)\,P(g_2/g_1) \qquad \text{(a)}$$
$$P(g_1 g_2) = P(g_2 g_1) = P(g_2)\,P(g_1/g_2) \qquad \text{(b)}$$
$$P(g_2)\,P(g_1/g_2) = P(g_1)\,P(g_2/g_1) \qquad \text{(c)}$$
$$P(g_2) = P(g_1)\frac{P(g_2/g_1)}{P(g_1/g_2)}. \qquad \text{(d)}$$

The ratio of $P(g_2/g_1)$ to $P(g_1/g_2)$ can be taken as a measure of the similarity of g_2 to g_1. The initial value of $P(g_2)$ depends on the amount of learning that has taken place in g_1, i.e., $P(g_1)$, and the degree of similarity of g_2 to g_1. The fact that the formula for similarity is a ratio of the kind that it is dramatizes the fact that similarity with respect to learning is a directional relationship, e.g., the amount of positive transfer from French to Spanish is not necessarily the same as the amount of positive transfer from Spanish to French.

Although the initial value of $P(g_2)$ is dependent linearly on $P(g_1)$ and the similarity of g_2 to g_1, it should be emphasized that the *increment* in learning is not a linear function of $P(g_1)$. Rather, the relationship is that as shown in Fig. 1.

The point of maximum learning for one reinforcement as a function of P_{n-1} varies with R. When R is high, the maximum is at low values of P_{n-1}. When R is low, the maximum is at higher values of P_{n-1}, although it is always at a value of P_{n-1} which is less than .50. Specifically, $P_n - P_{n-1}$ is maximal when [7]

$$P_{n-1} = \frac{\sqrt{R} - 1}{R - 1}. \qquad [5]$$

Curves of Learning

For the purposes of showing the progress in learning, formula [2] can be generalized to the form

$$P_n = \frac{R_1 R_2 \cdots R_n P_0}{R_1 R_2 \cdots R_n P_0 + (1 - P_0)}, \qquad [6]$$

where R_1 corresponds to x_1, etc., and P_0 is the probability of g prior to the verification of x_1.

For purposes of simplicity of exposition, let us assume that $R_1 = R_2 = \ldots = R_n$. The formula for the learning curve as a function of the number of reinforcements then becomes

$$P_n = \frac{R^n P_0}{R^n P_0 + (1 - P_0)}. \qquad [7]$$

Figure 2 then shows various learning curves for various values of R. These curves run from S-shaped to growth-type curves to curves of sharp rise which may be called insight curves. When R is low, we have "trial-and-error" learning, i.e., trial-and-error learning can be defined as learning which takes place when $P(x/gh)$ is not much larger than $P(x/\bar{g}h)$. Similarly, "insightful" learning takes place when $P(x/gh)$ is considerably larger than $P(x/\bar{g}h)$.[8]

Extinction

By analogous reasoning, it is possible to derive the phenomenon of extinction. Extinction is the consequence of the verification of $\bar{x}$. The result on the probability of g of the verification of $\bar{x}_m$ is

$$P_m = \frac{P_{m-1}}{P_{m-1} + \bar{R}(1 - P_{m-1})}, \qquad [8]$$

where m is the number of extinction trials, P_m is the probability of g after

[7] This was obtained by differentiating $P_n - P_{n-1}$ with respect to P_{n-1}, setting the derivative equal to zero, and solving for P_{n-1}. It has a limit of .50 as R approaches 1, R being 1 or greater. When $R = 1$, the maximum is indeterminate.

[8] Cf. Harlow: "The very form of the learning curve changes as learning sets become more efficient. The form of the learning curve for the first eight discrimination problems appears S-shaped: it could be described as a curve of 'trial-and-error' learning. The curve for the last 56 problems approaches linearity after trial 2. Curves of similar form have been described as indicators of 'insightful' learning" (5, p. 53).

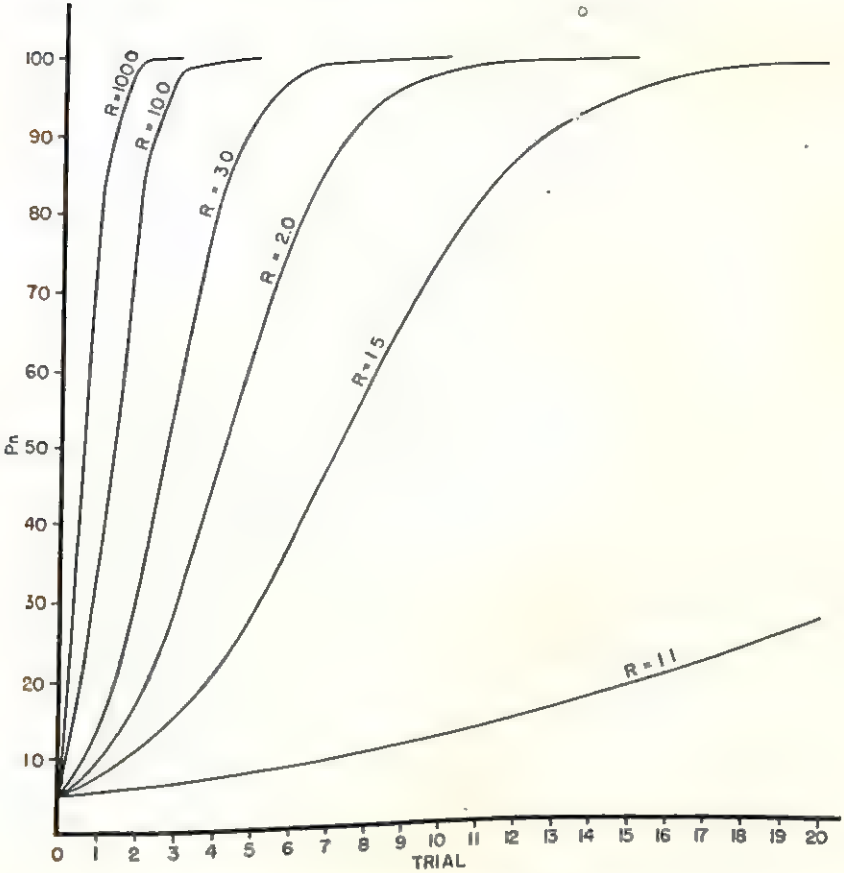


FIG. 2. Learning under a constant R

the verification of $\bar{x}_m$, P_{m-1} is the probability of g prior to the verification of $\bar{x}_m$, and $\bar{R}$ is the ratio of $P(\bar{x}/\bar{g}h)$ to $P(\bar{x}/gh)$. Figure 3 shows the effect of one negative reinforcement, the verification of $\bar{x}$ (x being found false). The equation for P_m, the probability of g after m successive extinction trials, is

$$P_m = \frac{P_0}{P_0 + \bar{R}_1 \bar{R}_2 \cdots \bar{R}_m (1 - P_0)}. \quad [9]$$

Figure 4 shows the extinction curves for constant values of $\bar{R}$.

DEDUCTIONS

It is beyond the scope of the present paper to elaborate on all the deductions that may stem from the theory of learning derived from the principle of inverse probability. However, in order to demonstrate its cogency, one example will be given.

What is the relation between the rate of learning and the rate of extinction? We have seen that if R is high the rate of learning is high, and that if $\bar{R}$ is high the rate of extinction is high. The question becomes one of determining the relationship between R and $\bar{R}$.

$$R = \frac{P(x/gh)}{P(x/\bar{g}h)}, \quad [10]$$

$$\bar{R} = \frac{P(\bar{x}/\bar{g}h)}{P(\bar{x}/gh)}. \quad [11]$$

Therefore, since $P(x/gh) + P(\bar{x}/gh) = 1$, and $P(x/\bar{g}h) + P(\bar{x}/\bar{g}h) = 1$,

$$\bar{R} = \frac{1 - P(x/\bar{g}h)}{1 - P(x/gh)}, \qquad [12]$$

thus fixing the relationship between R and $\bar{R}$. In order to complete the derivation, it is necessary to make an assumption. This assumption is that as we go from individual to individual, $P(x/gh)$ and $P(x/\bar{g}h)$ are positively correlated. This assumption is nothing more than the assumption that persons vary either in their supply of g's, or in their ability to draw x's from the g's that they have. If we make this assumption, then it follows immediately that R and $\bar{R}$ are negatively related, as can be seen from the following table:

Subject	$P(x/gh)$	$P(x/\bar{g}h)$	R	$\bar{R}$
1	.95	.80	1.19	4.00
2	.80	.65	1.23	1.75
3	.65	.50	1.30	1.43
4	.50	.35	1.43	1.30
5	.35	.20	1.75	1.23
6	.20	.05	4.00	1.19

The fact of the negative correlation between rate of learning and rate of extinction has been verified in a number of instances in conditioning studies (**2**, **3**, **10**, **14**).[9] It is also verified by the finding of the Gestalt psychologists that learning which has taken place by insight (R high) is not forgotten

[9] These studies are summarized by Hilgard and Marquis (**7**, p. 119).

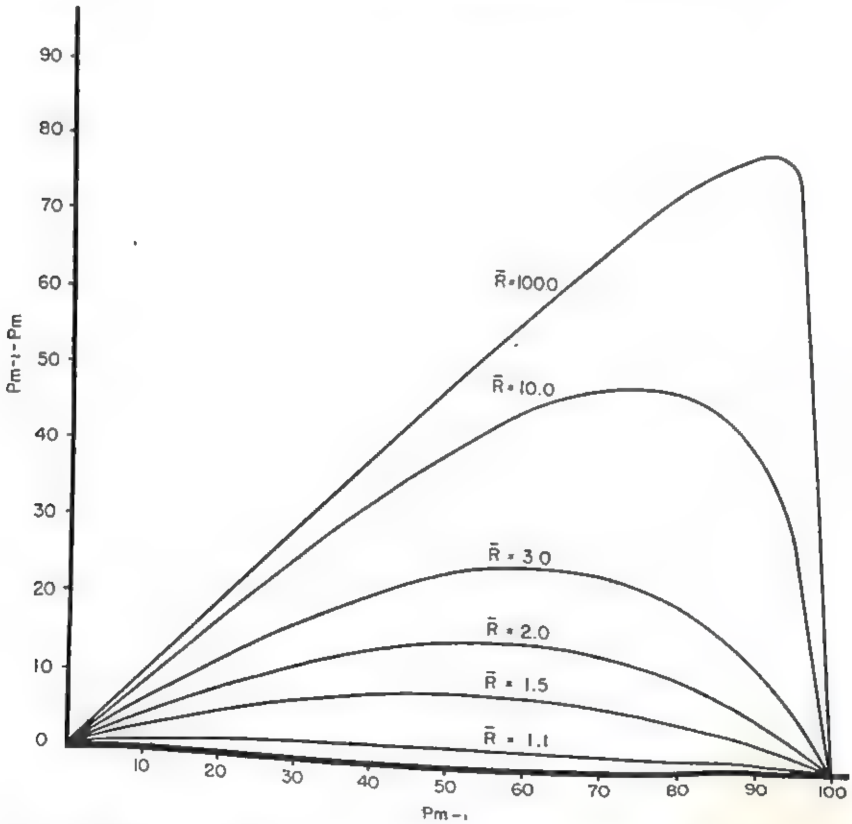


FIG. 3. The amount of decrement, $P_{m-1} - P_m$, as a function of one extinction trial, as a function of P_{m-1} and $\bar{R}$

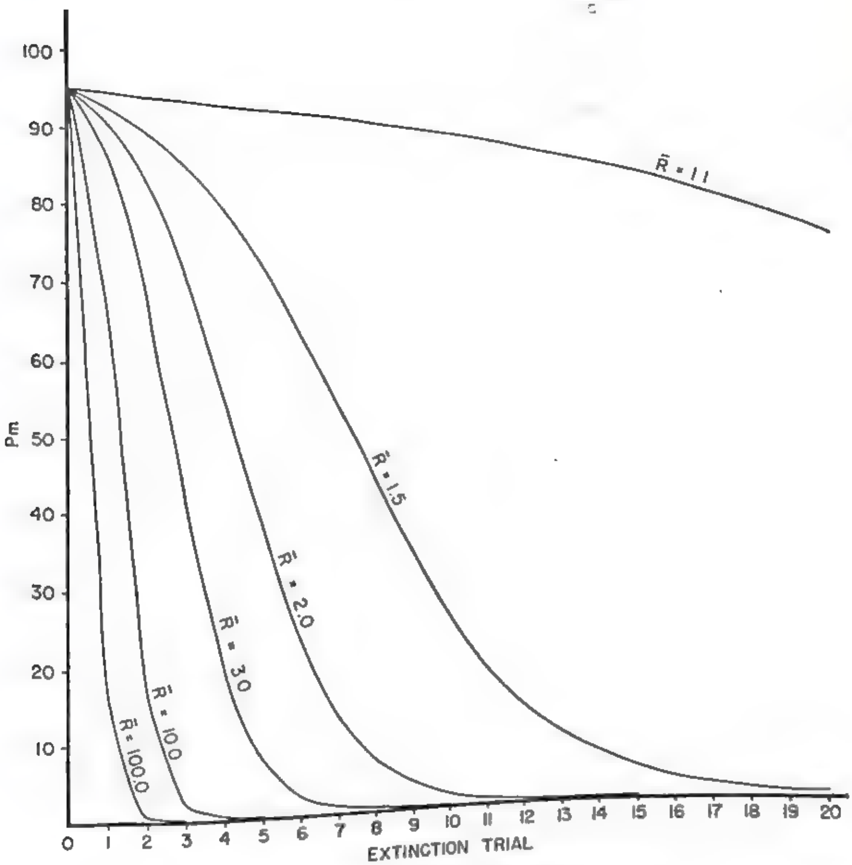


FIG. 4. Extinction under a constant $\bar{R}$

as quickly as learning which has taken place by rote (R low) (**12**), if we can, at least for the present, equate forgetting with extinction.

THEORETICAL IMPLICATIONS

As has been indicated, the primary purpose of this paper is simply to show the possibility of developing a theory of learning on the basis of the principle of inverse probability. However, taking the principle of inverse probability as a theory of learning has a number of implications with respect to psychological theory.

The method of science as a way of learning. In the opinion of the writer, the most important implication inherent in what has already been said is that by viewing science as a method of learning it is possible to learn something concerning the nature of learning. Whether the principle of inverse probability fully stands up against further scrutiny and empirical data with the interpretations which we have given to g, h, and x is to be seen. The burden of this paper is primarily to show the possibilities which are inherent in viewing science as a way of learning. Parenthetically it should be pointed out that the very testing of what has been advanced in this paper will employ the principle of inverse probability although it may not employ the particular interpretations of g, h, and x which we have advanced.

The reconciliation of stimulus-response, expectancy, and Gestalt theories. Although, in a sense, this theory of learning which we derive from the principle of inverse probability is primarily an expectancy theory, it incorporates some of the most important aspects of stimulus-response theory and Gestalt theory. For one thing, the phenomenon critical for Gestalt theory of learning, the insight phenomenon, is readily yielded as a consequence of the principle of inverse probability. Furthermore, it specifies, in a manner which the Gestalt theorists have not yet specified, the condition under which insight takes place, i.e., R is high.

One of the major superiorities of stimulus-response theory over expectancy theory has been the relative docility of the stimulus-response point of view to mathematical quantification, primarily in the hands of Hull and his followers. With the kinds of interpretation of the values which have been developed within the context of probability theory as has been employed here, the mathematics of probability becomes readily available for the handling of psychological problems. From a mathematical point of view, there is also a distinct superiority in the kind of quantification which has been outlined here. The heart of Hull's theory is in the equation for the learning curve. This was arrived at primarily on the basis of the relatively arbitrary method of curve fitting. There is really no good a priori rationale for determining whether the curve of learning should be a logarithmic or exponential growth function, as indicated by Hull's own wavering in this matter. On the other hand, the curve of learning which has been advanced in this paper is *necessary* on the basis of the fundamental postulates of probability theory and the interpretation given. One other point of superiority of the learning curve advanced in this paper over that of Hull's exponential growth function is, so to speak, its intuitive propriety. Although the criterion of intuitive propriety is not a necessary one, it is certainly a desirable one. Equation [1] (or better, perhaps, the theorem of compound probability), upon which the remainder is based, formulates what we all, in a sense, "know" to be true. In this respect, Hull's exponential growth function falls considerably short.

Account is taken of effect of "belongingness" on learning. Thorndike (16) found that it was necessary to take account of the "belongingness" or "relevance" of the reinforcement. When there was belongingness, learning was more rapid than when there was no belongingness. The degree of belongingness is exactly what is expressed by the value of R. If the value of $P(x/gh)$ is large, and $P(x/\bar{g}h)$ is small, then the verification of x "belongs" to g, or is relevant to g, and consequently, as has been indicated, the rate of learning is high.

Consistency with the general nativism of other theories. According to Thorndike and others, what is learned has, so to speak, to exist, at least in small measure in the organism. Thus, in Thorndikean terms learning consists in the strengthening of *already existing connections.* If the connections do not exist, at least in small measure, then they cannot be reinforced, and learning cannot take place.

In this respect, the present theory has to assume the very same thing. If learning is the growth of the probability of g, then there must be an initial probability of g which is greater than zero. If the initial probability of g is zero, then there can be no increase in the probability of g [3].

The possibility of understanding "complex" learning phenomena. Hilgard (6), in his opening paragraph, suggests that a theory of learning ought to be able to account for such phenomena as "prejudice and bigotry and other learnings which lead to trouble instead of to a satisfactory solution of . . . problems," as well as how skills, preferences, tastes, and knowledge develop. In most learning theories, these "complex" learnings are looked upon as a kind of higher level problem which will be solved after we have solved the problems involved in less "complex" learning. In the theory outlined, it is not necessary to differentiate between lower and higher learnings. The theory applies equally well to the learning of a conditioned response or a prejudice. The issue of molarity versus molecularity vanishes, as does the question as to whether reasoning and problem solving ought or ought not to be considered categories under the general heading of learning. Nor is it necessary to introduce the distinction between primary and secondary reinforcement to explain other than very simple learning.

HILGARD'S SIX QUESTIONS

Hilgard (6) has formulated six questions about theories of learning which are appropriately asked whenever a "new," or new version of an "old," theory is advanced. The attempt will be made to answer these questions with respect to what has been advanced here.

1. "*What are the limits of learning?*" Although the theory does not provide a contentual answer to this question, it supplies a formal answer. Learning is a function of $P(g/h)$, $P(x/gh)$, and $P(x/gh)$. The limits of learning are therefore in terms of the limits of h, the generation of g's on the basis of h, the generation of x's on the basis of the g's and h's, and whatever is involved in the determination of whether an x is or is not to be "tried." (The latter word is borrowed from Hilgard (6) as he uses it in the concept of "provisional try.") Unless there are x's, and unless the occasion for the verification of x's occurs, learning will not take place.

2. "*What is the role of practice in learning?*" Practice consists in the confirmation and nonconfirmation of x's. Repetition which does not involve the testing of x's will not result in learning. The amount of practice and its quality, the latter defined as the attributes of x, and the associated value of R are determiners of the amount of learning.

3. "*How important are reward, punishment, or other motives in learning?*" They are important in so far as they provide a basis for verifying x's, whose X's involve reward, punishment, or other consequences.

4. "*What is the place of understanding and insight?*" Whether learning is "blind" or "insightful" is a matter of degree as reflected in the magnitude of R.

5. "*Does learning one thing help you learn something else?*" Yes, in so far as the former thing learned has affected the initial probability of g as indicated by formula [4].

6. "*What happens when we remember and when we forget?*" Thus far in our development of the theory, we have only one suggestion to make with respect to this question. It has been shown that on the basis of the theory it follows that there is an inverse relation between R, the rate of learning, and $\bar{R}$, the rate of extinction. Katona's (12) investigations have shown that if learning is based on principle (R high), there is less forgetting ($\bar{R}$ low). Similar confirmations of the inverse

relationship between R and $\bar{R}$ have been found in conditioning studies. Thus, according to this, we will remember best that which has been learned with R high.

REFERENCES

1. Bakan, D. Learning and the scientific enterprise. *Psychol. Rev.*, 1953, **60**, 45–49.
2. Bernstein, A. L. Temporal factors in the formation of conditioned eyelid reactions in human subjects. *J. gen. Psychol.*, 1934, **10**, 173–197.
3. Campbell, A. A. The interrelations of two measures of conditioning in man. *J. exp. Psychol.*, 1938, **22**, 225–243.
4. De Morgan, A. *An essay on probabilities.* London: Longman, 1849.
5. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
6. Hilgard, E. R. *Theories of learning.* New York: Appleton-Century, 1948.
7. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
8. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
9. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.
10. Hunter, W. S. Conditioning and extinction in the rat. *Brit. J. Psychol.*, 1935, **26**, 135–148.
11. Jeffries, H. *Scientific inference.* New York: Macmillan, 1931.
12. Katona, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.
13. Keynes, J. M. *A treatise on probability.* London: Macmillan, 1948.
14. Mateer, R. *Child behavior.* Boston: Badger, 1918.
15. Nagel, E. Principles of the theory of probability. *Int. Encycl. Unif. Sci.*, 1947, **1**, No. 6. Chicago: Univer. of Chicago Press.
16. Thorndike, E. L. *Psychology of wants, interests and attitudes.* New York: Appleton-Century, 1935.
17. Uspensky, J. V. *Introduction to mathematical probability.* New York: McGraw-Hill, 1937.

[MS. received October 7, 1952]

Psychological Review
Vol. 60, No. 6, 1953

SIMULTANEOUS AND SUCCESSIVE DISCRIMINATION

M. E. BITTERMAN AND JEROME WODINSKY

University of Texas

In a recent experiment by Weise and Bitterman (12) the relative difficulty of simultaneous and successive discrimination was studied. A four-unit apparatus of the alley-maze type was employed, with two small lamps mounted at each choice point. One group of rats was trained on the simultaneous problem (i.e., turn in one direction when the right-hand lamp is on and in the opposite direction when the left-hand lamp is on) and a second group was trained on the successive problem (i.e., turn in one direction when both lamps are on and in the opposite direction when both lamps are off). The first problem proved to be much more difficult than the second, a result of considerable theoretical interest.

As Weise and Bitterman noted, the influential theory of Spence (8) was unable to account for the fact that the successive problem could be learned at all. Furthermore, a logical extension of the theory in accordance with the Hullian principle of afferent neural interaction (4), while making an explanation of successive learning possible, would lead to the deduction that the successive problem should be more difficult than the simultaneous. The Weise-Bitterman paper led Spence to modify his theory in the manner anticipated (9).[1] At the same time he reported the results of a new experiment with a simple elevated T maze (gray stem and black or white arms) in which the simultaneous problem was found to be considerably easier than the successive. Spence emphasized the correspondence of the new results with his extended theory and dismissed the contradictory data in an offhand manner: "Just why Weise and Bitterman got opposite results is not clear, as it is difficult to interpret the very complex type of discrimination set-up they employed. The simple discrimination situation is sufficiently difficult to deal with theoretically without adding all of the problems that arise as the result of the serial nature of the multiple discrimination set-up along with the fact that it involves a gradient of reinforcement" (9, p. 91).

Weise and Bitterman noted also the bearing of their data on the validity of Nissen's (5) attempt to subsume all discriminative behavior under the general headings of approach and avoidance. Nissen had deduced that the successive problem should be more difficult than the simultaneous problem since the former requires the development of conditional reactions or stimulus-compounding, a deduction which seemed to follow logically from a literal approach-avoidance theory. In a subsequent commentary on the Weise-Bitterman experiment, however, Nissen (6) denied that its results had any bearing on the validity of his formulation and attributed his deduction to the "misplacement" of a sentence. From this paper it became clear that Nissen's conceptual scheme was so broad as to be incapable of experimental evaluation (1).[2] Nissen did,

[1] Successive discrimination was explained in terms of compounding and was assumed, therefore, to be more difficult than simultaneous discrimination. For an evaluation of the extended theory, see Bitterman (2).

[2] The test which Nissen himself performed (5) is deprived of conclusiveness by his analysis of the Weise-Bitterman experiment. Animals were taught a simultaneous white-black discrimination with the stimuli horizontally

nevertheless, make one important criticism of the Weise-Bitterman experiment. He suggested that the less rapid learning of the simultaneous problem might have been due to reduction of the bright-dark difference by reflected light. From this point of view, Spence's contradictory results might be attributed to his use of painted stimuli rather than to the simplicity of his method.

These considerations led us to perform further experiments on the relative difficulty of simultaneous and successive problems with the conventional jumping apparatus. That situation should be simple enough to meet the requirements of Spence's theory, and the use of painted stimulus cards should avoid the criticism of the Weise-Bitterman apparatus which Nissen proposed. An as yet unpublished doctoral dissertation of E. F. MacCaslin provided evidence in support of the prediction by Weise and Bitterman that the relative difficulty of the two problems would depend on the similarity of the stimuli to be discriminated;[3] however, while the successive problem proved to be much more difficult than the simultaneous problem when difficult discriminations were employed (e.g., two vertically striped cards differing in stripe-width), even with simple discriminations (e.g., horizontally vs. vertically striped cards) the simultaneous problem was somewhat easier than the successive. In a subsequent experiment by Bitterman, Calvin, and Elam (3) with a discrimination between two circles markedly different in diameter, successive and simultaneous groups performed in almost identical fashion,[4] but in no experiment did we find superior performance in the successive group.

These results led us to reconsider the earlier interpretation of the Weise-Bitterman data. Another look at the maze apparatus made Nisson's explanation in terms of reflected light seem unlikely, and Spence's comment on complexity did not seem to further our understanding of the contradictory results—why the use of a multiple-discrimination apparatus should affect the relative difficulty of the two types of problem was not clear. Upon further consideration, however, another interpretation of the divergent results occurred to us. Weise and Bitterman assumed that their successive problem had been mastered on a configurational basis—its relative simplicity led them to reject the idea of compounding or conditional discrimination in favor of the assumption that the animals had learned merely to make one response to the bright configuration and an opposed response to the dark—but they expressed some doubt as to how the simultaneous problem had been solved: "If learned configurationally, the greater difficulty of the simultaneous problems may be attributed to the greater similarity between the two stimulus-patterns

arranged and later tested with a vertical arrangement of the same stimuli. Had there even been zero transfer, Nissen could have dealt with the results by assuming that the animals had learned to "approach" and "avoid" brightness-position compounds which were disrupted by the shift in the locations of the stimuli.

[3] Nissen was unkind enough to suggest that this prediction was made in anticipation of "the possibility of experimental evidence inconsistent with" the position of Weise and Bitterman (12, p. 164). Actually, the prediction followed logically from the data of Saldanha and Bitterman (7) on relational learning.

[4] In this experiment the simultaneous group had previously learned a simultaneous problem (horizontal vs. vertical stripes) while the successive group had been trained on a corresponding successive problem. For naive animals trained by MacCaslin on the circle discrimination, the simultaneous problem was somewhat easier. These studies point to the operation of qualitatively distinct perceptual sets in the two types of problem.

which it presented to the animal. . . .[5] If mastery was based on the acquisition of functional properties by afferent components, this kind of learning may be assumed to involve a more complex, higher order process" (12, pp. 192–193). Although Weise and Bitterman were led to emphasize the second interpretation, the experiment of Spence and our later studies with the jumping apparatus required us to re-examine the first.

In the jumping apparatus the animals were required to jump directly at the stimulus cards, and in Spence's T maze the animals were required to enter upon the stimulus runways. These may be described as *approach* situations in the literal meaning of the word. The relative simplicity of the simultaneous problem in such situations can be explained on the assumption that they facilitate the functional isolation of the two members of each pair of stimuli. In the apparatus of Weise and Bitterman, on the other hand, the two stimuli (lamps) at each choice point were closely juxtaposed and the animals were required to turn away from them, to one side or the other, in making their way through the maze. This may be described as a *response* situation. It facilitates configurational organization and thereby retards the functional isolation of the two members of each pair of stimuli. From this point of view the greater difficulty of the simultaneous problem is understandable either in terms of the greater similarity of its two configurations (on the assumption that solution is based on configurational discrimination) or in terms of the difficulty of perceptual analysis in such situations (on the assumption that solution is based ultimately on response to components). This interpretation, like Weise and Bitterman's, proposes two qualitatively distinct types of perceptual organization in these problems, but the functional priority of configurational perception is not postulated. Instead, it is assumed that the dominance of one or the other kind of organization is determined by the physical characteristics of the apparatus employed.

The experiment to be reported was designed to test the hypothesis that the Weise-Bitterman results were due to the fact that their apparatus—presenting closely juxtaposed stimuli which the animals were not required to approach directly—favored configurational organization. To distinguish between this interpretation and those of Spence and Nissen, it was necessary only to reproduce the essential features of the Weise-Bitterman situation in a suitably modified jumping apparatus.[6]

METHOD

Thirty experimentally naive Albino rats bred in the laboratory were studied. They were between three and four months old at the start of the experiment. After adjustment to handling, the animals were placed on a 24-hr. feeding schedule and preliminary training in the apparatus was begun.

The jumping apparatus was designed in the conventional way with only one exception. The two windows to which the animals were to jump were separated by a center window which was used only for stimulus cards (Fig. 1). The three windows were cut in a hemihexagonal surround which was painted gray. Behind the windows was a feeding platform to which the animals gained access following a correct response, and below the windows

[5] It was this statement that Nissen (6) apparently confused with his own analysis in terms of reflected light. Weise and Bitterman were merely noting the obvious fact that the configurations *dark-light* and *light-dark* are more similar than the configurations *dark-dark and light-light*. That this difference is independent of the stray-light problem the present experiment (with painted stimuli) will demonstrate.

[6] We are indebted to Mr. Richard Gonzalez for assistance in the conduct of this study.

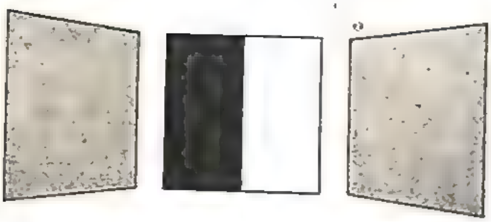


Fig. 1. Rat's view of the stimulus card (center) and response windows. The stimulus card shown was used in the simultaneous problem, in which the animal's task was to jump to one of the gray cards when the white area was on the right and to the other gray card when the white area was on the left. In the successive problem the stimulus card was either all white or all black, the white card signalling response to one of the windows and the black card to the other. The three windows in which the stimulus and response cards were set were cut in a hemihexagonal surround.

was a cloth net into which the animals fell after an incorrect response.

During preliminary training a gray card was locked in the center window, and the animals were trained to jump gradually increasing distances (up to a maximum of 9 in.) through the open left and right windows. Then they were trained to jump to unfastened gray cards (somewhat darker than the gray of the background to facilitate localization) in the left and right windows. Manual guidance was used to ensure equal experience with both windows. Following the preliminary training the animals were divided into two groups of 15 animals each which were matched for adjustment to the situation.

Group I was trained on the simultaneous problem and Group II on the successive problem. For both groups the gray cards used in the preliminary training appeared in the lateral windows on each trial. The center window, which contained the stimuli to be discriminated, was locked at all times as before. Group I was trained with a stimulus card which was half black and half white (Fig. 1). Seven of the animals in this group were rewarded for jumping to the right window when the white half of the card was on the right (black-white configuration) and to the left window when the white area was on the left (white-black configuration). The direction of correct response was reversed for the other eight animals of Group I (left to black-white and right to white-black). Group II was trained with two stimulus cards, both halves of each being either black or white. Seven of the animals were rewarded for jumping right to white-white and left to black-black, while the direction of correct response was reversed for the remaining eight animals of Group II (left to white-white and right to black-black). Ten trials per day, five to each of the two configurations of each problem, were given by the correction method. Each animal was allowed a maximum of three free jumps on each trial, and after three successive errors it was manually guided in the correct direction; a correct jump terminated each trial. The criterion of learning was one errorless day.

Results and Discussion

The course of learning in the two groups is plotted in Fig. 2 (initial errors) and Fig. 3 (total errors). In Table 1 the results are summarized in terms of mean errors and days to criterion. The difference between the two groups was large and statistically significant for all measures. The direction of the difference was in accord with that found by Weise and Bitterman—the performance of the successive group being markedly superior to that of the simultaneous group—and there was as little overlap between the two groups in the present experiment as in the earlier

TABLE 1

Relative Difficulty of Simultaneous and Successive Problems

	I Simultaneous	II Successive	*Diff.**
Days	14.0	7.9	6.1
Initial errors	50.7	25.2	25.5
Total errors	80.9	40.1	40.8

* All differences were significant beyond the 1% level of confidence by Wilcoxon's nonparametric test for unpaired deviates (13).

one. Thirteen of our successive animals reached the criterion of learning before the first simultaneous animal had done so, while in the earlier experiment 9 of the 10 successive animals reached criterion before the first simultaneous animal.

Although the present experiment confirms the results of Weise and Bitterman, accumulated evidence requires a new interpretation of those results. As we have seen, Weise and Bitterman assumed that their simultaneous problem represented a simple within-pairs differentiation, and on this assumption the greater simplicity of the successive problem was taken as evidence for the relatively primitive nature (defined in terms of priority or dominance) of configurational perception. It now seems likely that *both* the Weise-Bitterman problems were configurationally organized, at least to begin with, and that the difficulty of the simultaneous problem was due to the greater similarity between the configurations of that problem or to the fact that those configurations could not be readily differentiated into components. In situations such as Spence's T

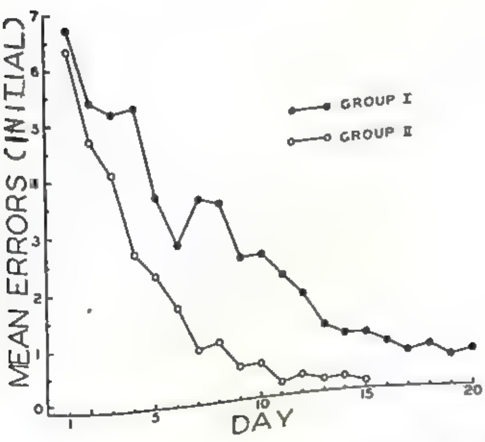


FIG. 2. The course of learning in the two groups plotted in terms of mean initial errors per day. Group I learned the simultaneous problem and Group II the successive problem.

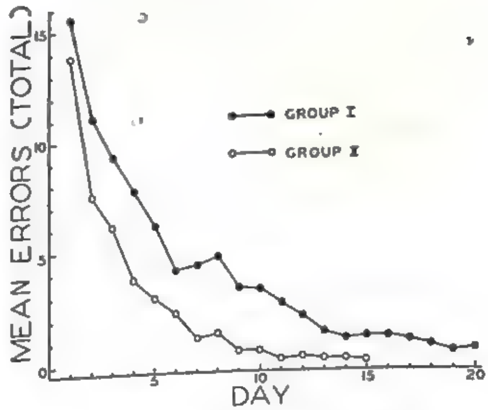


FIG. 3. The course of learning in the two groups plotted in terms of mean total errors per day. Group I learned the simultaneous problem and Group II the successive problem.

maze or the conventional jumping apparatus, which require the animals to approach (jump at or enter upon) the stimuli more directly, internal differentiation of simultaneous configurations is more readily achieved, and the simultaneous problem seems to be less difficult than the successive. Configurational effects may, nevertheless, operate in such situations (3, 14), and under specialized conditions, such as "two-situational" problems in which between-pairs differences are great and within-pairs differences are small, configurational organization may predominate (10, 11). At the present time the emphasis of Weise and Bitterman on qualitatively distinct levels of perceptual organization seems to be justified; unfortunately, however, the precise nature of these processes, their interrelations, and their hierarchical arrangement on a scale of priority remain in considerable doubt.

When the results of Spence and those obtained with the conventional jumping apparatus are contrasted with the results of Weise and Bitterman and those obtained in the present experiment, the value of a more restricted definition of the concept of *approach* is indicated.

If this term is used so loosely as to apply to behavior in all discriminative situations, no basis is provided for understanding the contrary results obtained in the two groups of experiments. As long as an animal is locomoting, it can be regarded as "approaching" something in its path and "avoiding" something not in its path;[7] such designations at best contribute nothing to our understanding and at worst may carry misleading implications concerning perceptual-motor relationships. If we look toward a more literal definition of approach-avoidance situations (perhaps in terms of the consequences of direct contact with stimulus objects), we have a basis for distinguishing at least two different classes of problem which seem to produce divergent results.

SUMMARY

The Weise-Bitterman experiment on the relative difficulty of simultaneous and successive discrimination has been criticized for complexity of method and failure to control stray light. In the present experiment, which was designed to forestall these criticisms, the Weise-Bitterman results were reproduced with a jumping apparatus. These and other recent experiments suggest the following conclusions: (*a*) in apparatus which require the animal to approach directly (jump at or enter upon) the stimuli to be discriminated, within-pairs differentiation is facilitated and the simultaneous problem is less difficult than the successive; (*b*) in apparatus in which the stimuli are closely juxtaposed and need not be directly approached, configurational organization is facilitated and the simultaneous problem becomes more difficult than the successive; (*c*) the qualitative distinction between relational (within-pairs) and configurational (between-pairs) discrimination continues to be applicable although the question of functional priority cannot be answered in general terms; (*d*) if the concept of approach-avoidance is to be useful, its definition must be considerably restricted.

[7] Nissen (5) has even extended the meaning of the terms to cover the appearance and non-appearance of "tail-twitching."

REFERENCES

1. BITTERMAN, M. E. Approach and avoidance in discriminative learning. *Psychol. Rev.*, 1952, 59, 172–175.
2. BITTERMAN, M. E. Spence on the problem of patterning. *Psychol. Rev.*, 1953, 60, 123–126.
3. BITTERMAN, M. E., CALVIN, A. D., & ELAM, C. B. Perceptual differentiation in the course of nondifferential reinforcement. *J. comp. physiol. Psychol.*, 1953, 46, 393–397.
4. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
5. NISSEN, H. W. Description of the learned response in discrimination behavior. *Psychol. Rev.*, 1950, 57, 121–131.
6. NISSEN, H. W. Further comment on approach-avoidance as categories of response. *Psychol. Rev.*, 1952, 59, 161–167.
7. SALDANHA, E. L., & BITTERMAN, M. E. Relational learning in the rat. *Amer. J. Psychol.*, 1951, 64, 37–53.
8. SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, 43, 427–449.
9. SPENCE, K. W. The nature of response in discrimination learning. *Psychol. Rev.*, 1952, 59, 89–93.
10. TEAS, D. C., & BITTERMAN, M. E. Perceptual organization in the rat. *Psychol. Rev.*, 1952, 59, 130–140.
11. TURBEVILLE, J. R., CALVIN, A. D., & BITTERMAN, M. E. Relational and configurational learning in the rat. *Amer. J. Psychol.*, 1952, 65, 424–433.
12. WEISE, P., & BITTERMAN, M. E. Response-selection in discriminative learning. *Psychol. Rev.*, 1951, 58, 185–195.
13. WILCOXON, F. *Some rapid approximate statistical procedures.* Stamford, Conn.: American Cyanamid Co., 1949.
14. WODINSKY, J., & BITTERMAN, M. E. Compound and configuration in successive discrimination. *Amer. J. Psychol.*, 1952, 65, 563–572.

[MS. received November 1, 1952]

Psychological Review
Vol. 60, No. 6, 1953

FORMALIZATION OF PSYCHOLOGICAL THEORY

F. J. McGUIGAN

Human Resources Research Office, George Washington University

Even though psychologists have been collecting data for over a century, the sizable amount of observational information which they now possess is in a relatively unsystematized state. It is particularly within the last two decades that psychologists have turned to the task of ordering these data. They have come to rather general agreement on the value of the hypothetico-deductive approach to accomplish this task, the first step of which is to choose or build an appropriate formal system. It is with this problem that the present paper is concerned. While we recognize several possible formal systems that can be utilized, it is difficult to say which can most adequately be interpreted in psychology. Only by trial of the more promising ones can an answer be found. Therefore, it behooves us to examine these possibilities and attempt to delimit this domain as well as possible in order that the more promising formal systems may receive the major part of our system construction efforts. This paper, then, will attempt to bring to bear on this problem the knowledge that is now available.

Nature of Formal Systems

A formal system is essentially a set of primitive terms combined to form postulates, from which additional terms and theorems can be derived. It is used in conjunction with (*a*) a set of formation rules which tell under what circumstances a symbol or combination of symbols may be considered to be a permissible expression, e.g., a proposition, and (*b*) a set of transformation rules which state the conditions for operations of deduction.

One type of logical expression with which we are particularly concerned is the synthetic formula. We use this type of expression for the statement of natural laws since it has the unique characteristic of being either true or false, depending on the empirical truth value of substitution instances of the variables employed. It is the prime goal of science to determine which specialized synthetic formulas are true and which are false. One rather widely accepted type of expression of natural laws is the general implication which allows prediction of a second variable, given the first.[1]

Once a scientist chooses a particular formal system to work with he draws from it a set of symbols and relates them to empirical referents. (The formal system is then said to be interpreted.) These symbols are then arranged to form a permissible expression, particularly a synthetic formula, in accordance with the appropriate formation rules. In the process of applying the formal system the scientist must assume that the synthetic statements with which he is working are true. It is not the function of formal systems, and the rules of formation and transformation, to determine which are true and which are false, but their function is exhausted by allowing derivation of new implications from the given synthetic formulas. It is rather the role of statistics to aid in determining the validity of synthetic formulas by determining the relative frequency of true to false instances.

[1] Reichenbach points out some of the difficulties incumbent in this type of expression for natural laws, and the possible solutions to them (9).

Natural Language

Psychology is now primarily in the stage of stating hypotheses in everyday language, and psychologists are becoming increasingly aware of the limitations of this procedure. It is a healthy commonplace to point out some of the defects of the scientific usage of natural language which can largely be overcome by using any formal system here discussed:

1. It generates ambiguity of the meaning of terms and hypotheses.
2. It does not allow easy translation of concepts between different theories.
3. It does not assure that implications of a theory follow in a rigorous fashion.
4. It provides only subjective criteria for deciding when a proposition is meaningful.
5. It does not sufficiently make accessible all of the implications contained in a hypothesis.
6. It makes impossible the evaluation of a theory according to the standard criteria of completeness, independence, and consistency which, while now not important, will become so when psychology reaches a more advanced stage of development.

Let us, then, briefly examine some of the more promising formal systems.

Logical Systems

Calculus of propositions. The primitive terms of this calculus include: (*a*) logical operations, "$-$," "$\vee$," "$\cdot$," "$\supset$," "$\equiv$," characterized respectively by the words *not, or, and, if . . . then,* and *material equivalence,* and (*b*) propositional variables—symbols which do not stand for any particular proposition but which occupy a place that can be filled by any special proposition and usually characterized by the letters *a*, *b*, etc.

In order to apply this formal system to psychology the elementary empirical propositions with which the theory is concerned should be coordinated to propositional variables. These elementary propositions, so symbolized, may then be combined to form more complex ones, in conjunction with the logical operations, as the theory maker desires. Such propositions may then be transformed according to the appropriate rules of transformation in order to arrive at deductions. This procedure may be illustrated by specializing the prototype for the expression of natural laws in Table 1 for the calculus of propositions. Thus, if we let *a* stand for the proposition "an animal is hungry" and *b* for "that animal will eat" and join these symbols by the logical operation of material implication, we have the statement of an extremely elementary law which can be interpreted as: "*If* an animal is hungry, *then* that animal will eat."

Calculus of functions. Whereas in the calculus of propositions the propositions are treated as wholes, in this calculus their inner structure is analyzed. The primitive terms here include: (*a*) term variables which hold open a place for things, usually characterized by the letters *x, y,* etc.; (*b*) predicate variables which hold open a place for properties, usually characterized by the letters *f, g,* etc.; (*c*) the logical operations mentioned in the calculus of propositions.

In applying this calculus the scientist must coordinate the things and properties with which his theory deals to either term or predicate variables and combine them in accordance with his formation

TABLE 1

Logical Expressions of Natural Laws

Calculus	Formula	Natural Language Interpretation
Propositions	$a \supset b$	"If *a* is true, then *b* is true"
Functions	$(x)[f(x) \supset g(x)]$	"For all *x* if *x* is *f*, then *x* is *g*"
Probability	$P[A, B] = p$	"If *A* is true, then *B* is probable to the degree *p*"

rules to form propositions. Taking the same simple example as previously used, and again referring to the appropriate prototype for expression of natural laws in Table 1, we will let f stand for "a hungry animal," and g for "will eat." The interpretation of this law in the vernacular would then be: "For all x, if x is a hungry animal, then x will eat."

Calculus of probability. Statements are made in this calculus by joining a known and an unknown class with the use of the probability implication, symbolized as "$\underset{p}{\ni}$." This term is the only new primitive term not found in the calculi previously discussed and is interpreted in the context of the basic abbreviated formula of this calculus, "$a \underset{p}{\ni} b$," as "if a is true, then b is probable to the degree p." In order to facilitate applications of this calculus, solutions for the degree p, etc., the probability implication is abbreviated by using equations of the type indicated in Table 1. It is then possible to utilize the general form of the previous calculi discussed within the "P symbol," e.g., the calculus of functions, thus allowing us to perform the same transformations as in these other calculi. An illustration of our example using this calculus would be to let A stand for "An animal is hungry," and B stand for "that animal will eat"; then we can arrive at the following interpretation: "The probability from the statement 'An animal is hungry' to 'That animal will eat' is equal to p." The value of p can then be determined by computing the relative number of true to total instances of empirical observations.

An additional characteristic of this calculus is its ability to handle numerically quantified data in addition to qualitative data. This is done by enumerating elements of a class that possess numerical properties instead of qualitative properties. For an elaboration of this procedure see Reichenbach (10).

One problem that is made particularly apparent in the application of this calculus concerns the assumption that the sequence of events being studied has a limit of frequency at p. This is supposedly a necessary condition for an admissible interpretation, and yet the calculus must be applied prior to knowing that this assumption is satisfied. The general answer to this problem, as Reichenbach (10) points out, is that we do not know for sure that we have chosen a sequence that so converges, but the only chance for success is to proceed as if it does. The criterion of our success is the adequacy of our formulation for making successful predictions. More particularly, we obtain the initial section of a sequence and assume that the value of p thus calculated will persist, within certain limits of exactness, for the rest of the sequence.

Evaluation of logical systems. Since the calculi of propositions and functions are two-valued systems of logic, i.e., the specialized variables contained in the propositions stated in them must be either completely true or completely false, they suffer various disadvantages. Thus, if a given type of behavior is a function of several factors acting simultaneously, the amount contributed by these factors must be taken into account in order to yield a prediction, yet this cannot be done with these calculi. Similarly, predictions cannot be made about a precise quantitative value of an unknown variable, as can be done in, e.g., an algebraic equation. The calculus of functions, when compared to the calculus of propositions, is generally more fertile for yielding deductions since it allows internal analysis of propositions, instead of treating propositions as wholes. The calculus of probability, when utilizing the calculus of functions

within the "P symbol," however, is generally more fertile than either of the other two calculi since it is capable of handling both qualitative and numerically quantitative data. Thus, not only can this calculus handle amounts of variables entering into a behavioral equation, but it can also be used to predict numerically quantitative values of unknown variables.

The two-valued systems of logic are limited in another very important respect. Thus, if any given empirical proposition is subjected to test and but one instance of negation occurs, regardless of the number of positive instances observed, the proposition should properly be considered false. There is no way to state the relationship of positive to negative instances in these calculi. All synthetic statements, however, have a probability character, which makes the calculus of probability the only known logical system rigorously appropriate for application to the physical world. By its use, probability values can be attached to propositions in terms of positive and negative instances observed.

One particular advantage of any of these calculi over the more classical mathematical systems should be pointed out. That is that they all may be used to perform rigorous transformations of qualitatively stated propositions.

Mathematical Systems

While it is rather widely accepted that mathematics is reducible to logic we are making a working distinction between these two disciplines. Further, within the field of mathematics we choose to make an arbitrary distinction between classical mathematical systems and unique mathematical systems at a working level.

Classical mathematical systems. This category is intended to include systems which deal with numerically quantified data such as algebra, trigonometry, integral and differential calculus, etc., the primitive terms of which are mathematical variables (x, y, etc.), mathematical constants (a, b, etc.), and various operations such as addition, subtraction, etc. In utilizing these systems the scientist can avail himself of either of two procedures, depending on the state of knowledge in his area. The first consists of discovering relationships from empirically plotted data and fitting the appropriate mathematical functions to those data. In this case there need not be a specific hypothesis in the sense that the parameters of the equation need not have predetermined interpretations. The scientist may then speculate about the parameters of his equation and coordinate empirical terms to them. If such be the case, these interpreted equations would then represent hypotheses which should be validated by testing them against new data.

The second procedure consists of defining empirical terms, coordinating them to formal symbols, and combining these symbols according to the formation rules appropriate to the system involved. Deductions may then be arrived at by putting the equations through the various mathematical operations, e.g., differentiating an algebraic equation. These equations then represent explicit hypotheses which may be subjected to test by fitting them to appropriate empirical data and determining their goodness of fit. This involves the general equation used and the generality of the mathematical constants involved. It is hoped that both are very general in that they do not vary to any great extent for a variety of experimental situations. It is likely, however, that this generality will be limited, e.g., while the general form of the equation may hold throughout a variety of situations, the constants might be limited by specific

types of learning tasks, measurements taken, experimental situations, etc.

Unique mathematical systems. Mathematics has largely developed according to the needs of one science, physics. It is these particular types of mathematics, i.e., classical mathematics, that scientists are most familiar with, and that, therefore, are assumed to be the proper formal systems to use in psychology. There is no particular reason, however, to think that these types of mathematics are particularly applicable in psychology, and it is likely that new systems will have to be developed or discovered. There are a number of instances in the history of science where empirical sciences were able to make great advances only after singularly appropriate formal systems were made available.

Lewin (6) represents the outstanding attempt to apply a unique mathematical system to psychology, which is essentially a composite of vector geometry, topology, and classical mathematics. He attempted to develop empirical constructs which characterize the life space, coordinate them to topological and vector concepts, and then state the laws governing changes in these properties, e.g., he coordinated a need to the vector concept of a system in tension. One unique characteristic of his attempts at formalization is that he utilized formal systems that could handle qualitative data. The answer to the question concerning adequacy of these formal systems is not unequivocal since considerably more work should be done towards this end. Actually Lewin did not progress much farther in applying these geometries than to use a few surface concepts, but what has been done, e.g., his formalization of Zeigarnik's studies, appears to be a definite advancement.

Other preliminary attempts in this direction that appear promising are the efforts of Estes (3) and Bush and Mosteller (1, 2) to utilize set theory in conjunction with classical mathematics.

DISCUSSION

We have indicated that two-valued logic and classical mathematics, taken separately, are inadequate systems for psychology. There thus appear to be three more promising possibilities for attempts at formalization: (*a*) a combination of two-valued logic and classical mathematics, (*b*) a combination of classical mathematics and a unique type of mathematics that is capable of handling qualitative data, e.g., topology, and (*c*) the calculus of probability.

However, in view of the fact that our laws should take the form of a probability statement, the only really appropriate system seems to be probability logic, since it has this capacity, in addition to all the apparent advantages of the other systems. It is, of course, possible that this calculus can be used in combination with a mathematical system, but the conclusion that at least it should be used is inescapable, and it is likely that it alone can be sufficient. There is, on the other hand, no reason to think that only one formal system will be appropriate in psychology but it may well turn out that, as in physics, several geometries can be interpreted with equal adequacy. However, since probability logic is somewhat more complicated to use than two-valued logic and since it is possible to interpret the statements of one in the other without any great difficulty, it may be that the latter can be used to arrive at approximations.

There are, of course, other logical systems that could have been discussed, e.g., calculus of classes, three-valued logic. However, these possess no apparent advantages over the calculi here presented. Thus, both the calculus of

classes and three-valued logic are generalized in probability logic, and neither offers apparent advantages as techniques for approximations over the calculus of propositions or the calculus of functions.

Some examples of two-valued systems applied in psychology are the works of Hull *et al.* (5), Miller (7), and Fitch and Barry (4). Woodger applies them in biology (11), and Reichenbach (8), in the area of quantum mechanics, makes the first application of three-valued logic.

A note of caution is now in order. We must be aware of the waste of energy that would come with premature formalization; this does not mean, however, that we are not ready for it any place in psychology. We should be discriminating and attempt formalization only in well-developed areas where its success is a real possibility. One problem in formalization is the difficulty of encompassing all the well-established facts in a field. This is another indication that our initial attempts should be made in limited and well-founded fields, e.g., aspects of learning, and not attempt to formalize the whole field of psychology at this time. Through experience we will learn to what extent our limited interpreted systems can be concatenated to provide a unified whole in psychology.

Formalization is probably the major problem facing us today. Our more advanced sciences have shown us that great advances come with increased formalization. This is the trend—it is on us. Furthermore, psychologists are the ones who will have to do most of the work. We cannot rely on logicians and mathematicians, because they do not have the necessary psychological knowledge with which to work, and if they did, psychologists would still have to have a working knowledge of the formal system in order to grasp the real meaning of propositions. Interpretations back into the vernacular fail to convey the full essence of complex hypotheses which are formally stated.

REFERENCES

1. Bush, R. B., & Mosteller, F. A mathematical model for simple learning. *Psychol. Rev.*, 1951, **58**, 313–323.
2. Bush, R. B., & Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413–423.
3. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
4. Fitch, F. B., & Barry, Gladys. Towards a formalization of Hull's behavior theory. *Phil. Sci.*, 1950, **17**, 260–265.
5. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven, Conn.: Yale Univer. Press, 1940.
6. Lewin, K. *Field theory in social science.* New York: Harper, 1951.
7. Miller, J. G. Symbolic technique in psychological theory. *Psychol. Rev.*, 1939, **46**, 464–479.
8. Reichenbach, H. *Philosophic foundations of quantum mechanics.* Berkeley and Los Angeles: Univer. of California Press, 1946.
9. Reichenbach, H. *Elements of symbolic logic.* New York: Macmillan, 1947.
10. Reichenbach, H. *The theory of probability.* Berkeley and Los Angeles: Univer. of California Press, 1949.
11. Woodger, J. H. The technique of theory construction. *Int. Encycl. Unif. Sci.*, 2, No. 5. Chicago: Univer. of Chicago Press, 1947.

[MS. received **November 8, 1952**]

Psychological Review
Vol. 60, No. 6, 1953

INHIBITION THEORY AND THE EFFORT VARIABLE [1]

DOUGLAS S. ELLIS

Iowa State College

The present paper analyzes a motor learning problem which seems to be somewhat confused. At the theoretical level, the problem is concerned with the Hullian constructs of reactive inhibition and conditioned inhibition (12, pp. 277–303). At the empirical level, it consists of determining functional relationships between the effortfulness of the task and certain measures of task performance.

The significance of the effort-inhibition problem stems from the theoretical orientation of much recent motor learning research. A good deal of this research has been concerned with testing hypotheses deduced from Hull's inhibitory constructs, and has shown that the effects of a number of variables on motor task performance are consistent with expectations from inhibition theory. However, such a satisfying state of affairs does not appear to exist for the effort variable. Several *E*s have made predictions concerning the effects of effort on inhibition (8, 10, 11). Data obtained by these and other *E*s have failed to support the predictions (4, 5, 8, 10, 11).

The thesis of the present paper is that this effort-inhibition research has been misinterpreted, largely because of difficulties inherent in the theory. Following a brief statement of the basic operations and concepts involved in such research, this thesis will be developed by considering three topics: (*a*) the inhibitory constructs, (*b*) the deductions which can properly be made from them, and (*c*) the implications of past research in the light of these more proper deductions. Finally, suggestions for future research will be advanced.

Basic Operations and Concepts

Operations. Tasks used in Hullian-oriented motor learning research have included reversed alphabet printing, the pursuit rotor, a block-turning task, a ball-placing task, and a hand-cranking task. The experimental treatment usually accorded *S* is the performance of such a task before and after a single interpolated rest. The basic response measure has been *S*'s rate of motor performance, although the major interest has been in response measures derived from this simple measure (see below). The variables manipulated have included number of prerest responses, degree of distribution of prerest practice, length of interpolated rest, and, of course, effort.

The *E*s have manipulated effort in a number of ways. In the pursuit rotor, Eckstrand attached vertical springs of varying tension to the rotor stylus. The *S* had to work against this spring tension in order to bring the stylus down to the plane of the revolving target (8). Helmick, who also used the pursuit rotor, believed that he was manipulating effort in varying the speed of rotation of the target (11). In the block-turning

[1] The author wishes to express his thanks to Drs. Benton J. Underwood and Carl P. Duncan, of Northwestern University, and to Richard B. McHugh, of Iowa State College, for their helpful comments on the manuscript. Thanks are also due to George W. McNelly for the many hours spent in discussing issues involved in this paper. Some of these issues were discussed in a paper, jointly authored with Mr. McNelly, presented at the 1952 meeting of the Midwestern Psychological Association.

task, an attempt was made to manipulate effort by varying the height of the work surface on which the task rested (10). From previous research (9) it was assumed that high work-surface heights were more effortful than moderate heights. More direct manipulations have been used by Bilodeau in simple motor tasks. In a ball-placing task, balls of varying weights were used (4). In a crank-turning situation, Bilodeau varied the force necessary to turn the crank (5).

Theory. These experimental operations have been largely dictated by the desire to test hypotheses deduced from inhibition theory. The basic notions of the theory are simple. It is essentially, as Kimble has noted (15), a two-factor theory of inhibition. Two inhibitory processes are postulated; both are presumed to depress performance. However, the two kinds of inhibition are differentiated in respect to their permanence. One of these inhibitory processes, reactive inhibition (I_R), is temporary in the sense that it dissipates with time. The other, conditioned inhibition ($_SI_R$), is permanent in that it is a habit, a learned response of not-responding. ($_SI_R$ is of course not absolutely permanent; like other habits, it can be forgotten or unlearned.) These two inhibitory factors are presumed to summate to yield a total inhibitory potential, which operates to depress performance.

Indices of the inhibitory constructs. Such a theory yields deductions couched in inhibition language. However, one does not measure inhibition, one measures motor behavior. Thus it is necessary to determine derived response measures which will serve as indices of I_R and $_SI_R$. Both Ammons (1) and Kimble (15) have suggested such indices. Reminiscence, taken as the extent to which a group's initial postrest performance exceeds its final prerest performance, has been used as an index of I_R. The notion is that reminiscence is due to the dissipation of the I_R generated prior to rest, and that the amount of reminiscence thus reflects the amount of I_R dissipated over rest. The index for $_SI_R$ is more complicated. It is based on the notion of comparing initial postrest performance with the performance expected if prerest practice had been free from both kinds of inhibition. A distributed-practice condition which has received approximately the same amount of practice is customarily used to provide the estimate of inhibition-free performance. Figure 1 illustrates this technique of measuring I_R and $_SI_R$ in an experimental group. The distance A is reminiscence, the I_R index. The distance B, which represents the extent to which the experimental group fails to recover to the level of the distributed group, is the $_SI_R$ index.

Past research. These are the two response measures that have been of major research interest. Of the two, the I_R index has received the most attention. It has been shown, as would be expected from theory, that reminiscence is (*a*) directly proportional to the number of prerest responses (2, 4, 10, 13, 15, 18), (*b*) in-

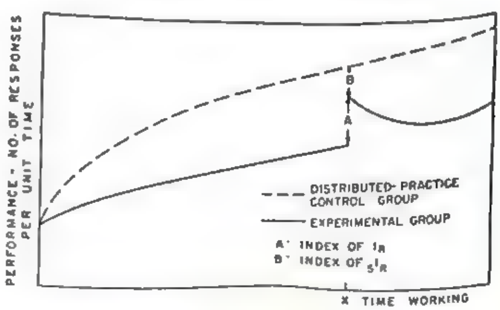


FIG. 1. Indices of reactive inhibition (I_R) and conditioned inhibition ($_SI_R$). A rest is interpolated at time X for the experimental group. Indices noted estimate the amount of I_R and $_SI_R$ generated prior to rest in the experimental group.

versely proportional to the degree of distribution of prerest practice (3, 6, 16), and (*c*) directly proportional to the length of the interpolated rest (2, 4, 10, 13, 14).

Results from the few experiments using the $_SI_R$ index present a less clear picture. Substantial amounts of $_SI_R$, which behave as expected from theory, have been found in two studies (15, 20). However, a recent study (19) suggests that this measured $_SI_R$ may be an artifact of experimental procedure, and several other studies (3, 16, 18) find little evidence for $_SI_R$. We shall carry the $_SI_R$ construct in our theoretical analysis, but it should be recognized that the evidence for such a construct is, at best, fragmentary.

As has been indicated previously, results from the effort variable do not appear to fit neatly with the theory. The usual prediction is that the I_R index (reminiscence) is proportional to effort. The usual finding is that the I_R index is independent of effort (4, 5, 8, 10, 11). No research has been published to date concerning the effects of effort on the $_SI_R$ index. In a subsequent section it will be shown that this prediction of a proportionality between I_R and effort does not necessarily follow from Hull's inhibition theory.

THE INHIBITORY CONSTRUCTS

Reactive inhibition. In his theoretical formulation, Hull specifies three characteristics of his constructs: (*a*) their psychological status, (*b*) the manipulable stimulus variables which influence them, and (*c*) the response from which they may be inferred. I_R is given motivational status; it is described as a response-avoidance drive. The variables which are explicitly stated to determine the amount of I_R generated by a sequence of responses are two: (*a*) the number of responses evoked, and (*b*) the work involved in the execution of the response. The time elapsed since the cessation of responding is a variable of special importance for I_R. I_R is given the special property of spontaneously dissipating with time. Finally, I_R is presumed to reflect itself in behavior through its postulated characteristic of decreasing the probability that the response will occur.

The relationship between I_R and work specified above deserves further comment on two counts. The first is a matter of terminology. In the present paper we are using the term "effort" to describe essentially what Hull means by "work." A number of considerations dictated this choice. First, after defining work as the product of forces times length, Hull points out that the inhibition associated with a given amount of work is not merely a matter of mechanics, but is partially a function of the strength of the muscle system involved (12, p. 279). Thus, predictions must concern themselves with more than just the gross physical work accomplished by a response. Second, work can denote several things: physical work (foot-pounds), the fact of performing a set of duties (I went to work), the repeated execution of a well-practiced response (19, p. 555), a subjective experience (this is hard work). Third, the trend among current authors seems to be to employ effort rather than work.

It is of course possible to effect a transposition between effort, as used here, and work, defined as the product of force and length. Most *Es*, although they have described their manipulations in terms of effortfulness, have actually endeavored to manipulate work by varying the force requirements of the task. Eckstrand's variation of spring tension in the pursuit rotor, or Bilodeau's variation of the weight of the balls used in the

ball-placing task are clear examples of this. The work-surface height manipulation is not so direct. For this variable one might speculate that the muscular tension generated at high work-surface heights represents an opposing force which has to be overcome in responding. Helmick's variation of the rate of rotation of the pursuit-rotor target is a manipulation which attempts to vary work by varying length, since the rate increase lengthens the target track per unit time. Actually, both the work-surface height and rotor-rate variables are somewhat suspect as work manipulations. For the work-surface height variable it would have to be shown that variation does influence the force requirements of responding. For the rotor-rate variable, it would be necessary to show that the force exerted is equivalent at the rate values used.

The second general comment refers to the fact that the I_R-work relationship given above is actually a simplification of Hull's theorizing. The equation given by Hull is (**12**, p. 279):

$$\text{(i)} \qquad \dot{I}_R = \frac{cn}{B - W},$$

where c and B are constants, n is the number of responses evoked, and W is work.

Confusion attends the $\dot{I}_R$ term. In later portions of the section on inhibition, $\dot{I}_R$ is defined as the total inhibitory potential according to the equation (**12**, p. 285):

$$\text{(ii)} \qquad \dot{I}_R = I_R + {}_sI_R.$$

If equation (i) is taken as stands, then it allows no separation of the effects of W on I_R and ${}_sI_R$. W could influence either or both of the two inhibitory components.

The simplification lies in assuming that Hull was referring to I_R, not $\dot{I}_R$, in equation (i). There are two reasons for such an assumption. First, the equation occurs in the section discussing the characteristics of I_R. Second, this simplification has been made implicitly by a number of previous authors. For example, Kimble, who was one of the first to apply Hullian inhibitory concepts to motor learning problems, has stated that I_R is proportional to work or effort (**14**, p. 239; **16**, p. 500).

Conditioned inhibition. ${}_sI_R$ is given the status of a habit. The reinforcement for this habit is presumed to be the dissipation of I_R which accompanies the cessation of work. When work ceases, I_R dissipates and provides the reinforcing state of affairs necessary to condition the cease-work response to stimuli present. The stimulus variables which influence ${}_sI_R$ are not explicitly stated. However, it seems apparent that the amount of ${}_sI_R$ developed will be primarily a direct function of the amount of I_R present and the rapidity with which it dissipates. Thus, the amount of ${}_sI_R$ generated will hinge on the variables previously discussed which determine the amount of I_R present. Finally, like I_R, ${}_sI_R$ is presumed to decrease the probability of recurrence of the response evoked.

Deduction of the Effects of Effort on the Inhibitory Constructs

Reactive inhibition. It would seem, from the theory sketched above, that an experimental test of the proportionality between I_R and work or task effort would be straightforward. Work could be varied by altering the effortful demands of the task, and I_R could be indexed by the amount of reminiscence shown after the introduction of an interpolated rest. The theoretical expectation would be that I_R and its reminiscence index would be proportional to work. Unfortu-

nately, the theoretical expectation is not strictly correct. The proportionality between I_R and work holds only for a special case.

Let us examine this theoretical difficulty. It stems from the fact that there is another variable besides n (number of responses) and W (work) which determines the amount of I_R generated by a sequence of responses. This variable, rate of responding (r) is *implicit* in the theoretical formulations of Hull.[3] To illustrate the role of r, consider two response sequences which are equivalent in n and W, but which have different r's. If one sequence is evoked at a very slow rate, it would be possible for all the I_R generated by a response to dissipate before the next response occurred. In this case, the amount of I_R remaining at termination would be merely the amount generated by the last or terminal response. This situation can be contrasted with the case where responses are evoked at a rapid rate. Here, I_R could accumulate throughout the sequence, since there would not be sufficient time for it to dissipate appreciably between responses. In this case the amount of I_R present at termination would be the amount generated by the terminal responses *plus* the undissipated I_R remaining from prior responses.

With the rate variable explicit, the theory takes on quite a different appearance. In essence, it states the following three functional relationships:

(iii) $$I_R = g(n)_{Wr}$$

(iv) $$I_R = h(W)_{nr}$$

(v) $$I_R = i(r)_{nW}$$

[3] In fairness to Kimble, it should be pointed out that he has given the rate variable explicit status (14, p. 239). However, Kimble has not dealt with the effort variable as such, and thus the rate variable apparently has remained implicit for most Es working on the effort-inhibition problem.

where g, h, and i denote functional relationships (I_R increasing with increases in n, W, or r), the subscripts indicate variables which are held constant, and I_R refers to amount of reactive inhibition generated by the response sequence.

While these equations are useful when one is content to vary one response sequence characteristic at a time, additional equations would be needed to handle cases where more than one characteristic varied. Equation (i) is such an additional equation; it handles cases where both n and W vary but r remains constant. However, a final relationship, specifying either the functional relationship of n to W or r to W is not given. Thus, the theory does not handle the general case. It does not permit predictions concerning the relative amount of I_R generated by sequences which differ in W, n, and r. Some of the effort-inhibition research has been of this nature. The Ss have worked for a given period of time at one of several conditions of effortfulness. The effortful manipulation has depressed rate of responding and thus also altered the number of responses evoked (**4**, **5**, **11**).

However, although it does not follow for the general case, the usual prediction of a proportionality between I_R and work or effort does hold for a special case. From equation (iv) it can be seen that the proportionality follows from theory if response sequences varying in W, but identical in n and r, are compared. This would occur, for example, if the effortful manipulation did not depress rate of responding, and if the groups of Ss at the various effort conditions all worked for the same length of time. Some of the effort-inhibition research has been of this nature (**8**, **10**). For example, in the work-surface height study (**10**), Ss working at high (presumably more effortful) heights re-

sponded just as rapidly prior to rest as *S*s who worked at moderate heights. Since both groups worked for the same length of time prior to rest they were equivalent in total number of prerest responses, and the study thus compares response sequences equivalent in *n* and *r*, but supposedly different in effortfulness or work. It will be noted subsequently (see implications of past research) that, while the failure of the effortful manipulation to depress rate of responding rescues the predictions of such experiments, it also raises some doubt as to the adequacy of the effortful manipulation.

Conditioned inhibition. The comments regarding the inability of inhibition theory to handle the I_R-effort function for the general case also have implications for the $_SI_R$-effort function. Since $_SI_R$ depends for its development on the dissipation of I_R, the theory does not permit predictions in the general case concerning the effects of effort on $_SI_R$.

DEDUCTIONS FROM MODIFIED HULLIAN THEORY

Although the theory, as it stands, is indeterminate, it can be coupled with a supporting assumption to yield predictions. This supporting assumption can be seen most clearly within the framework of Kimble's extension of Hull's theory. Kimble adds the notion that organisms will only tolerate a threshold or critical amount of I_R, and that when this threshold is reached the organism rests. For Kimble, this rest may actually occur during the response sequence. Here is Kimble's picture of an organism working under relatively massed practice (15, p. 16):

> Since I_R is a drive, it seems only reasonable to suppose that the accumulation of a certain critical amount will automatically produce resting. . . . Presumably, once I_R is reduced to below the critical level, the organism driven by motivation to perform the task at hand will resume work and continue working until the critical level of I_R is reached again. Then it will rest, reducing I_R; start work again, increasing I_R and so on.

Now the supporting assumption is merely that this critical level of I_R is the same for all conditions of effort. Thus, in an effort-reminiscence experiment, the assumption would be that *S*s in the various effort conditions would tolerate approximately equal amounts of I_R. In terms of Kimble's theory, which assumes (16, p. 500), with some empirical basis (17, 20), that motivation is the primary determinant of the critical I_R level, this amounts to assuming that the *S*s are equally motivated at the various effort conditions. Or, in alternative terms, one might characterize the assumption as stating that *S*s in the various effort conditions are allotting the same amount of energy to the task at hand. Intuitively, this seems a reasonable assumption. It also has the virtue, as will be seen subsequently, of yielding predictions which are in accord with experimental results.

Reactive inhibition. Figure 2 sketches the development of I_R at two conditions of effort under the modified theory. Consider the curve for the

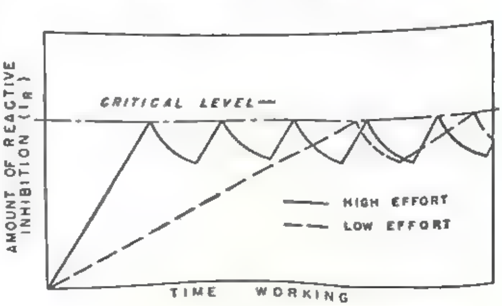


FIG. 2. Theoretical generation of reactive inhibition (I_R) at two conditions of task effort. The curves shown assume a slower rate of responding at the high effort condition than at the low effort condition.

low effort condition. As responses are evoked, I_R cumulates until the critical level is reached. When the amount cumulated reaches this level, the organism stops work, and I_R dissipates. After sufficient I_R has dissipated, the organism starts work again and the cycle is repeated.

The curve for the high effort condition portrays the same sequence of events. Since we have not assumed any particular relationship between task effort and rate of responding, the curve shown is just one of a family of possible high effort curves, each curve corresponding to a particular rate of responding. The particular curve shown rests on the assumption that rate of responding is somewhat depressed at the high effort condition. There are other possibilities. If the effort conditions were identical in rate of responding, the high effort curve would rise even more steeply. If high effort *S*s adjusted their rate of responding so that they were generating the same amount of I_R per unit time as the low effort *S*s, the high and low effort curves would coincide. Finally, if high effort *S*s responded so slowly that they were generating less I_R per unit time than low effort *S*s, the high effort curve would fall below the low effort curve.

The major point of the figure lies in what happens once the two conditions reach the critical level. Irrespective of just what rate of responding adjustment high effort *S*s have adopted, once the critical level is reached the effort conditions should be equivalent in I_R. The I_R-work prediction from the modified theory is thus: *after a moderate amount of practice, I_R is independent of effort.*

Some question naturally arises concerning what is meant by a moderate amount of practice. In his theory, Kimble points out that I_R probably rises quickly to a maximum. If reminiscence is taken as an index of I_R, then empirical amount practice-reminiscence functions furnish estimates of how much time is required for I_R to reach a maximum. Numerous studies (4, 10, 13, 14) show that the balance of the growth of I_R is accomplished in the first 3–5 min. practice.

Conditioned inhibition. Figure 2 also permits some comment concerning the effects of effort on $_SI_R$. It will be recalled that, in Kimble's theory, it is the resting responses pictured which develop $_SI_R$. Figure 2 shows that both effort conditions have identical amounts of I_R present when rest is taken. This, therefore, would not produce differential amounts of $_SI_R$ at the two conditions. However, Fig. 2 also points up the possibility that the conditions might differ in the number of resting responses made. This could produce a difference in $_SI_R$: the condition with the greater number of resting responses should develop the most $_SI_R$.

Whether or not the conditions will differ in number of resting responses and $_SI_R$ cannot be predicted from the theoretical formulation given, since it will depend on the kind of rate adjustment adopted by high effort *S*s. For example, if the adjustment shown in Fig. 2 is assumed, there would be more $_SI_R$ (more resting responses) at the high effort condition. An alternative assumption is that *S*s adjust their rate of responding to different values of effort so that they are generating comparable amounts of I_R per unit time. This assumption stretches our original equality-of-motivation assumption, which was restricted to points in practice after reaching the critical level, over the entire work period. It would place the curves of Fig. 2 together and yield the prediction that both I_R and $_SI_R$ are inde-

pendent of task effort at all stages of practice.[3]

IMPLICATIONS OF PAST RESEARCH

With predictions concerning the effects of effort on I_R and sI_R clarified, let us consider the results of the five relevant studies. At the outset it should be noted that none of these studies has been directly concerned with sI_R, so that the discussion will be limited to the I_R construct and its index, reminiscence. The studies may be divided into two groups: (*a*) those in which the effortful manipulation did not depress prerest performance, and (*b*) those in which the effortful manipulation depressed prerest performance.

The Eckstrand (8) and the Ellis, Montgomery, and Underwood (10) studies belong in the first group. Since prerest rate of responding was not depressed in either study, the correct prediction, according to the previous discussion (see deduction of task effort effects, special case), would be that I_R and reminiscence would be proportional to task effort. In contradiction to this theoretical expectation, both studies found that reminiscence was comparable at the various effortful conditions.

However, as has been pointed out (10), there is a consideration which may vitiate the theoretical implications of these studies. Since there appears to be no case in the literature where equal prerest performance is followed by unequal amounts of reminiscence, one might expect that effortful conditions with identical prerest performance would show identical reminiscence.

The Helmick (11) and Bilodeau (4, 5) studies belong to the second class of studies. In these studies the effortful manipulations did depress prerest performance. For this case, the prediction from our theoretical section would be that I_R and reminiscence would be independent of effort. This is essentially the finding of the studies. There is some tendency for reminiscence to increase slightly with increases in effort, but there is no statistical evidence that this trend is significant. Thus, it appears that results from contemporary effort-inhibition research are consistent with a slightly modified version of Hullian inhibition theory.

FUTURE RESEARCH

The remedy for those studies in which the effortful manipulation failed to depress prerest performance seems simple. Choose an effortful manipulation and a response measure so that the response measure is sensitive to the effortful manipulation. Actually, this remedy is not so simple. It appears difficult to find response measures which are sensitive to certain types of effortful manipulations. Consider the work-surface height variable. The *S*s working at heights approximately on a level with their shoulders emit signs of effort, such as profanity and requests to cease work, but persistently refuse to satisfy *E* with depressed performance. Or consider Eckstrand's variation of the spring tension on the pursuit-rotor stylus. At the greatest spring tension, evidence indicated that *S*s had difficulty in working for the required time. Yet, performance was not depressed under this condition.

These observations suggest that there may be two general classes of effortful manipulations, and that the

[3] Bilodeau, in handling the results of one of his studies, has recently arrived at a view generally similar to that presented in this section on deductions from modified Hullian theory. Bilodeau suggests that the *S*s in the various effortful conditions were about equally fatigued, and that they had achieved this state of affairs through adjustments in their rate of responding (5, p. 99).

kind of response measure which will be sensitive will depend on what class of effortful manipulation is being used. The two classes may be called response-dependent and response-independent effort.

Work-surface height and stylus spring tension appear to be good examples of response-independent effort because the effort is largely independent of response events. For work-surface height the feature which appears to generate effort is not turning blocks over, but holding arms up. Stylus spring tension has the same general characteristic: the effort exists independently of whether *S* is on or off target. Ordinary performance measures may not be sensitive to such response-independent effortful manipulations. What may be needed is some index of how long *S* will work voluntarily at the task (7). Studies of the influence of task effort on reminiscence scores based on such an index would determine whether the Hullian constructs are useful in handling phenomenon associated with response-independent effort.

As examples of response-dependent effort, Bilodeau's ball-weight and crank-force manipulations can be cited. Here the generation of effort is directly contingent on the evocation of the response measured. Variation of the weight of the blocks in the block-turning task is another example of response-dependent effort. Such a manipulation for the pursuit rotor is harder to envisage. Perhaps an apparatus which provided an increasing gradient of spring tension as the stylus approached the target would be the answer. Our point would be that, for such response-dependent effortful manipulations, ordinary performance measures would be appropriate.

The need for some principle of motivation or energy allotment, so apparent in the theoretical section of this paper, also suggests a fresh approach to some motor learning problems. Specifically, two types of problems may be suggested.

First, although our analysis suggests that one of the basic invariances in the effort experiment is the motivation or amount of energy which *S* allots to the task, we have no guarantee that this invariance will stand up at the extremes of the effort dimension. It may be that present research has been limited to effortful manipulations which cluster at the high or low end of the effort dimension, and that a more adequate exploration of the dimension would show up some systematic relationship between effortfulness and inhibition. Certainly it would appear that some research should be directed at determining features of situations which upset *S*s' usual methods of adjusting energy output.

A second type of problem involves the direct investigation of the brief inhibition-dissipating resting responses postulated by Kimble. The study of long sequences of responses evoked under massed practice should provide information as to the presence of these resting responses and their role in the mechanics of inhibition, and also suggest the laws which govern their appearance. The scheme of energy allotment which *S* achieves with these responses may well be a potent variable in motor learning situations.

SUMMARY

The present paper has dealt with the implications of effort-reminiscence research for Hullian inhibition theory. In the past, such research has been viewed as yielding results inconsistent with this theory.

It is pointed out here that this research has been misinterpreted. First, it is noted that Hullian theory, as stated, does not permit general predictions concerning the effects of

effort on reminiscence. A prediction is presented for the special case where effort does not depress performance, but it is pointed out that such an effortful manipulation does not permit critical tests of the theory. Second, it is shown that, if Hullian inhibition theory is coupled with supporting assumptions concerning how *S*s allot energy to the task at hand, a prediction of independence between reminiscence and effort can be made. Third, since this is the result obtained by past studies, it is concluded that past research on the reminiscence-effort function is indeed consistent with a modified form of Hullian inhibition theory. Finally, research designed to capitalize on the energy allotment problem made explicit in the theoretical section of the paper is suggested.

REFERENCES

1. AMMONS, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
2. AMMONS, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
3. AMMONS, R. B. Acquisition of motor skill: III. Effects of initially distributed practice on rotary pursuit performance. *J. exp. Psychol.*, 1950, **40**, 777–787.
4. BILODEAU, E. A. Performance decrement in a simple motor task before and after a single rest. *J. exp. Psychol.*, 1952, **43**, 381–390.
5. BILODEAU, E. A. Decrements and recovery from decrements in a simple work task with variation in force requirements at different stages of practice. *J. exp. Psychol.*, 1952, **44**, 96–100.
6. BILODEAU, E. A. Massing and spacing phenomena as a function of prolonged and extended practice. *J. exp. Psychol.*, 1952, **44**, 108–113.
7. BOLDT, R. F., & ELLIS, D. S. Number of responses and rate of responding required to reach a voluntary rest pause as a function of response effortfulness. Paper read at Midwest. Psychol. Ass., Cleveland, 1952.
8. ECKSTRAND, G. A. Post-rest performance in motor learning as a function of amount of work performed in pre-rest practice. Unpublished doctor's dissertation, Northwestern Univer., 1949.
9. ELLIS, D. S. Speed of manipulative performance as a function of work-surface height. *J. appl. Psychol.*, 1951, **35**, 289–296.
10. ELLIS, D. S., MONTGOMERY, V., & UNDERWOOD, B. J. Reminiscence in a manipulative task as a function of work-surface height, pre-rest practice, and interpolated rest. *J. exp. Psychol.*, 1952, **44**, 420–427.
11. HELMICK, J. S. Pursuit learning as affected by size of target and speed of rotation. *J. exp. Psychol.*, 1951, **41**, 126–138.
12. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
13. IRION, A. L. Reminiscence in pursuit-rotor learning as a function of rest and amount of pre-rest practice. *J. exp. Psychol.*, 1949, **39**, 492–499.
14. KIMBLE, G. A., & HORENSTEIN, B. R. Reminiscence in motor learning as a function of length of interpolated rest. *J. exp. Psychol.*, 1948, **38**, 239–244.
15. KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
16. KIMBLE, G. A. Performance and reminiscence in motor learning as a function of degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.
17. KIMBLE, G. A. Evidence for the role of motivation in determining the amount of reminiscence in pursuit rotor learning. *J. exp. Psychol.*, 1950, **40**, 248–253.
18. REYNOLDS, B., & BILODEAU, INA MCD. Acquisition and retention of three psychomotor tasks as a function of distribution of practice during acquisition. *J. exp. Psychol.*, 1952, **44**, 19–26.
19. SCHUCKER, R. E., STEVENS, L. B., & ELLIS, D. S. A retest for conditioned inhibition in the alphabet-printing task. *J. exp. Psychol.*, 1953, **46**, 97–102.
20. UNDERWOOD, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.
21. WASSERMAN, H. N. The effect of motivation and amount of pre-rest practice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, **42**, 162–172.

[MS. received November 21, 1952]

Psychological Review
Vol. 60, No. 6, 1953

AN APPROACH TO THE STUDY OF COMMUNICATIVE ACTS

THEODORE M. NEWCOMB

University of Michigan

This paper points toward the possibility that many of those phenomena of social behavior which have been somewhat loosely assembled under the label of "interaction" can be more adequately studied as communicative acts. It further points to the possibility that, just as the observable forms of certain solids are macroscopic outcomes of molecular structure, so certain observable group properties are predetermined by the conditions and consequences of communicative acts.

The initial assumption is that communication among humans performs the essential function of enabling two or more individuals to maintain simultaneous orientation toward one another as communicators *and* toward objects of communication. After presenting a rationale for this assumption, we shall attempt to show that a set of propositions derived from or consistent with it seems to be supported by empirical findings.

Co-orientation and the A-B-X System

Every communicative act is viewed as a transmission of information, consisting of discriminative stimuli, from a source to a recipient.[1] For present purposes it is assumed that the discriminative stimuli have a discriminable object as referent. Thus in the simplest possible communicative act one person (A) transmits information to another person (B) about something (X). Such an act is symbolized here as AtoBreX.

The term "orientation" is used as equivalent to "attitude" in its more inclusive sense of referring to both cathectic and cognitive tendencies. The phrase "simultaneous orientation" (hereinafter abbreviated to "co-orientation") itself represents an assumption; namely, that A's orientation toward B and toward X are interdependent. A-B-X is therefore regarded as constituting a system. That is, certain definable relationships between A and B, between A and X, and between B and X are all viewed as interdependent. For some purposes the system may be regarded as a phenomenal one within the life space of A or B, for other purposes as an "objective" system including all of the possible relationships as inferred from observations of A's and B's behavior. It is presumed that a given state of the system exists when a given instance of AtoBreX occurs, and that as a result of this occurrence the system undergoes some change (even though the change be regarded as only a reinforcement of the pre-existing state).

The minimal components of the A-B-X system, as schematically illustrated in Fig. 1, are as follows:

1. A's orientation toward X, including both attitude toward X as an object to be approached or avoided (characterized by sign and intensity) and cognitive attributes (beliefs and cognitive structuring).
2. A's orientations toward B, in exactly the same sense. (For purposes of avoiding confusing terms, we shall speak of posi-

[1] This statement is adapted from G. A. Miller's definition: "'information' is used to refer to *the occurrence of one out of a set of alternative discriminative stimuli.* A discriminative stimulus is a stimulus that is arbitrarily, symbolically, associated with some thing (or state, or event, or property) and that enables the stimulated organism to discriminate this thing from others" (9, p. 41).

tive and negative *attraction* toward A or B as persons, and of favorable and unfavorable *attitudes* toward X.)

3. B's orientation toward X.
4. B's orientation toward A.

In order to examine the possible relationships of similarity and difference between A and B, we shall make use of simple dichotomies in regard to these four relationships. That is, with respect to a given X at a given time, A and B will be regarded as cathectically alike (++ or −−) or different (+− or −+) in attitude and in attraction; and as cognitively alike or different. We shall also make use of simple dichotomies of degree—i.e., more alike, less alike. We shall refer to lateral similarities of A's and B's orientations to X as *symmetrical* relationships.

This very simple system is designed to fit two-person communication. In the following discussion these additional limitations will be imposed, for simplicity's sake: (*a*) communicative acts will be treated as verbal ones, in face-to-face situations; (*b*) initiation of the communicative act is considered to be intentional (i.e., such acts are excluded as those which the actor assumes to be unobserved); (*c*) it is assumed that the "message" is received—i.e., that the communicative act is attended to by an intended recipient, though not necessarily with any particular degree of accuracy; and (*d*) A and B are assumed to be group members, characterized by continued association.

The assumption that co-orientation is essential to human life is based upon two considerations of complementary nature. First, the orientation of any A toward any B (assuming that they are capable of verbal communication) is rarely, if ever, made in an environmental vacuum. Even in what seems the maximally "pure" case of two lovers oblivious to all but each other, both singly and both jointly are dependent upon a common environment; and their continued attachment is notoriously contingent upon the discovery or development of common interests beyond themselves. It is not certain that even their most person-oriented communications (e.g., "I love you") are devoid of environmental reference. The more intense one person's concern for another the more sensitive he is likely to be to the other's orientations to objects in the environment.

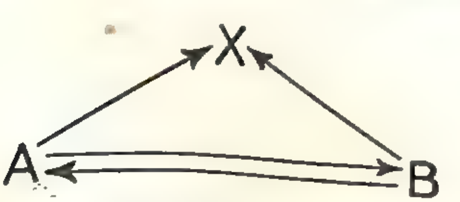


FIG. 1. Schematic illustration of the minimal A-B-X system

Second, the orientation of any A capable of verbal communication about almost any conceivable X is rarely, if ever, made in a social vacuum. There are few if any objects so private that one's orientations toward them are uninfluenced by others' orientations. This is particularly true with regard to what has been termed "social reality" (3); i.e., the less the possibility of testing one's assumptions by observing the physical consequences of those assumptions, the greater the reliance upon social confirmation as the test of what is true and valid. And even when assumptions can be put to the direct test (e.g., the child can find out for himself about the stove which he has been told is hot), social reality is often accepted as the quicker or the safer test. As various linguists have pointed out, moreover, a good deal of social reality is built into the very language with which we communicate about things. Under the conditions of continued association which we are assuming, A and B as they communicate about X are dependent

upon each other, not only because the other's eyes and ears provide an additional source of information about X, but also because the other's judgment provides a testing ground for social reality. And to be dependent upon the other, in so far as such dependence influences behavior, is to be oriented toward him.

In short, it is an almost constant human necessity to orient oneself toward objects in the environment and also toward other persons oriented toward those same objects. To the degree that A's orientation either toward X or toward B is contingent upon B's orientation toward X, A is motivated to influence and/or to inform himself about B's orientation toward X. Communication is the most common and usually the most effective means by which he does so.

SYMMETRY OF ORIENTATION

Much of the remainder of this paper will deal with the relationships between A's and B's orientations toward X, within the postulated A-B-X system. The implications of this model are: (*a*) that while at any given moment the system may be conceived of as being "at rest," it is characterized not by the absence but by the balance of forces; and (*b*) that a change in any part of the system (any of the four relationships portrayed in Fig. 1) may lead to changes in any of the others. We shall also make the assumption (not inherent in the model) that certain forces impinging upon the system are relatively strong and persistent, and that thus there are "strains" toward preferred states of equilibrium.

This assumption, related to the initial one concerning the co-orientation function of communication, is as follows. To the degree that A's orientation toward X is contingent upon B's orientation toward X, A's co-orientation will be facilitated by similarity of his own and B's orientation toward X. The first advantage of symmetry—particularly of cognitive symmetry—is that of ready calculability of the other's behavior; the more similar A's and B's cognitive orientations, the less the necessity for either of them to "translate" X in terms of the other's orientations, the less the likelihood of failure or error in such "translations," and thus the less difficult and/or the less erroneous the co-orientation of either. Second, there is the advantage of validation of one's own orientation toward X; the more similar A's and B's orientations, either cognitive or cathectic (particularly in the many areas where validation is heavily dependent upon "social reality"), the more confident each of them can be of his own cognitive and evaluative orientations. Co-orientation is of course possible with little or no symmetry, but the facilitative value of symmetry for co-orientation is considerable.

If these advantages are commonly experienced as such, communicative acts resulting in increased symmetry are likely to be rewarded, and symmetry is likely to acquire secondary reward value. This is the basis of our assumption of a persistent "strain toward symmetry," under the conditions noted.

These assumptions may now be brought together in terms of the following inclusive postulate: *The stronger the forces toward A's co-orientation in respect to B and X,* (*a*) *the greater A's strain toward symmetry with B in respect to X*; and (*b*) *the greater the likelihood of increased symmetry as a consequence of one or more communicative acts.* The latter part of the postulate assumes the possibility of modified orientations toward X on the part of both A and B, who over a period of time exchange roles as transmitters and receivers of information.

Several testable propositions are derivable from this postulate. First, if the likelihood of instigation to and achievement of symmetry varies as a function of forces toward co-orientation, the latter varies, presumably, with valence of the objects of co-orientation—i.e., of intensity of attitude toward X and of attraction toward B. That is, under conditions such that orientation toward either B or X also demands orientation toward the other, the greater the valence of B or of X the greater the induced force toward co-orientation, and thus the greater the likelihood of both instigation toward and achievement of symmetry.

Such research findings as are known to the writer are in support of these predictions. Experimental results reported by Festinger and Thibaut (5), by Schachter (12), and by Back (1) indicate that attempts to influence another toward one's own point of view vary as a function of attraction. In the second of these studies it is shown that communications within a cohesive group are directed most frequently toward those perceived as deviates, up to a point where the deviate is sociometrically rejected (i.e., attraction decreases or becomes negative), beyond which point communication to them becomes less frequent. It is also shown in this study that frequency of influence-attempting communication varies with degree of interest in the topic of group discussion.

Some of these same studies, and some others, present data concerning symmetry as a consequence of communication. Thus Festinger and Thibaut, varying "pressure toward uniformity" and "perception of homogeneous group composition," found actual change toward uniformity following a discussion to be a function of both these variables, but some change toward uniformity took place in every group, under all conditions. Back found that subjects who started with different interpretations of the same material and who were given an opportunity to discuss the matter were influenced by each other as a direct function of attraction.

Findings from two community studies may also be cited, as consistent with these laboratory studies. Newcomb (10), in a replicated study of friendship choices as related to political attitudes in a small college community, found on both occasions that students at each extreme of the attitude continuum tended to have as friends those like themselves in attitude. Festinger, Schachter, and Back (4), in their study of a housing project, found a correlation of +.72 between a measure of attraction and a measure of "conformity in attitude." No direct observations of communication are made in these two studies; the relevance of their findings for the present point depends upon the assumption that frequency of communication is a function of attraction. This assumption is clearly justified in these two particular investigations, since in both communities there was complete freedom of association. As noted below, this assumption is not justified in all situations.

Other testable propositions derivable from the general postulate have to do with A's judgments of existing symmetry between himself and B with respect to X. Such judgments (to which the writer has for some time applied the term "perceived consensus") are represented by the symbol *B-X*, within A's phenomenal A-B-X system. Such a judgment, under given conditions of demand for co-orientation with respect to a given B and a given X, is a major determinant of the likelihood of a given AtoBreX, since strain toward symmetry is influenced by perception of existing symmetry. Such a judgment, more-

over, is either confirmed or modified by whatever response B makes to AtoBreX. The continuity of an A-B-X system thus depends upon perceived consensus, which may be viewed either as an independent or as a dependent variable.

According to the previous proposition, the likelihood of increased symmetry (objectively observed) as a consequence of communicative acts increases with attraction and with intensity of attitude. The likelihood of perceived symmetry presumably increases with the same variables. Judgments of symmetry, like other judgments, are influenced both by "reality" and by "autistic" factors, both of which tend, as a function of attraction and intensity of attitude, to increase the likelihood of perceived consensus. Frequency of communication with B about X is the most important of the "reality" factors, and this, as we have seen, tends to vary with valence toward B and toward X. As for the "autistic" factors, the greater the positive attraction toward B and the more intense the attitude toward X, the greater the likelihood of cognitive distortion toward symmetry. Hypothetically, then, perceived symmetry with regard to X varies as a function of intensity of attitude toward X and of attraction toward B.

A considerable number of research studies, published and unpublished, are known to the writer in which subjects' own attitudes are related to their estimates of majority or modal position of specified groups. Only a minority of the studies systematically relate these judgments to attraction, and still fewer to intensity of attitude. Among this minority, however, the writer knows of no exceptions to the above proposition. The most striking of the known findings were obtained from students in several university classes in April of 1951, in a questionnaire dealing with the very recent dismissal of General MacArthur by President Truman:

	pro-Truman *Ss* who . . .	anti-Truman *Ss* who . . .
attribute to "most of my closest friends"		
pro-Truman attitudes	48	2
anti-Truman attitudes	0	34
neither	4	4
attribute to "most uninformed people"		
pro-Truman attitudes	6	13
anti-Truman attitudes	32	14
neither	14	13

If we assume that "closest friends" are more attractive to university students than "uninformed people," these data provide support for the attraction hypothesis. Comparisons of those whose own attitudes are more and less intense also provide support, though less strikingly, for the hypothesis concerning attitude intensity.

Perceived symmetry, viewed as an independent variable, is obviously a determinant of instigation to symmetry-directed communication. Festinger (3), with specific reference to groups characterized by "pressures toward uniformity," hypothesizes that "pressure on members to communicate to others in the group concerning item x increases monotonically with increase in the perceived discrepancy in opinion concerning item x among members of the group," as well as with "relevance of item x to the functioning of the group," and with "cohesiveness of the group." And, with reference to the choice of recipient for communications, "The force to communicate about item x to a particular member of the group will increase as the discrepancy in opinion

between that member and the communicator increases [and] will decrease to the extent that he is perceived as not a member of the group or to the extent that he is not wanted as a member of the group" (3, p. 8). Support for all of these hypotheses is to be found in one or more of his and his associates' studies. They are consistent with the following proposition: the likelihood of a symmetry-directed AtoBreX varies as a multiple function of perceived discrepancy (i.e., inversely with perceived symmetry), with valence toward B and with valence toward X.

Common sense and selected observations from everyday behavior may also be adduced in support of these propositions. For example, A observes that an attractive B differs with him on an important issue and seeks symmetry by trying to persuade B to his own point of view; or A seeks to reassure himself that B does not disagree with him; or A gives information to B about X or asks B for information about X. From all these acts we may infer perception of asymmetry and direction of communication toward symmetry. Selected observations concerning symmetry as a consequence of communication are equally plentiful; there is, in fact, no social phenomenon which can be more commonly observed than the tendency for freely communicating persons to resemble one another in orientation toward objects of common concern. The very nature of the communicative act as a transmission of information would, on a priori grounds alone, lead to the prediction of increased symmetry, since following the communication both A and B possess the information which was only A's before. B will not necessarily accept or believe all information transmitted by A, of course, but the likelihood of his doing so presumably varies not only with attraction toward A but also with intensity of attitude toward X, since in the long run the more important X is to him the more likely it is that he will avoid communicating with B about X if he cannot believe him. Thus the propositions have a considerable degree of face validity.

But everyday observation also provides instances to the contrary. Not all communications are directed toward symmetry, nor is symmetry an inevitable consequence of communication, even when attraction is strong and attitudes are intense. A devoted husband may refrain from discussing important business matters with his wife, or two close friends may "agree to disagree" in silence about matters of importance to both. People who are attracted toward one another often continue to communicate about subjects on which they continue to disagree—and this is particularly apt to happen with regard to attitudes which are intense, contrary to our theoretical prediction.

In sum, the available research findings and a considerable body of everyday observation support our predictions that instigation toward, perception of, and actual achievement of symmetry vary with intensity of attitude toward X and attraction toward B. The readiness with which exceptions can be adduced, however, indicates that these are not the only variables involved. The propositions, at best, rest upon the assumption of *ceteris paribus;* they cannot account for the fact that the probabilities of A's instigation to communicate about a given X are not the same for all potential B's of equal attraction for him, nor the fact that his instigation to communicate to a given B are not the same for all X's of equal valence to him. We shall therefore attempt to derive certain further propositions from our basic assumption that both instigation to and achievement of symmetry vary with strength of forces toward co-orientation in the given situation.

DYNAMICS OF CO-ORIENTATION

The foregoing propositions represent only a slight extrapolation of Heider's general principle (6) of "balanced states" in the absence of which "unit relations will be changed through action or through cognitive reorganization." In a later paper devoted specifically to the implications of Heider's hypotheses for interrelationships among attitudes toward a person and toward his acts, Horowitz *et al.* (8) note the following possible resolutions to states of imbalance: (*a*) the sign-valence of the act is changed to agree with that of the actor; (*b*) the reverse of this; and (*c*) the act is cognitively divorced from the actor; in addition, of course, the disharmony may be tolerated.

Orientations as attributed by A to B are here considered as equivalent to acts so attributed, in Heider's sense, and symmetry is taken as a special case of balance. Assume, for example, the following asymmetry in A's phenomenal system: +A:X, +A:B, $-B{:}X$, $+B{:}A$ (i.e., A has positive attitude toward X, positive attraction toward B, perceives B's attitude toward X as negative, and B's attraction toward A as positive). Any of the following attempts at "resolution," analogous to those mentioned by Heider, are possible: (*a*) −A:X; (*b*) −A:B; or (*c*) cognitive dissociation. These can occur in the absence of any communication with B. Attempts at harmony (symmetry) may also be made via communications directed toward $+B{:}X$. And, if such attempts fail, the three alternatives mentioned as possible without communication are still available. Finally, there is the possibility of compromise, following communication (e.g., agreement on some midpoint), and the possibility of "agreeing to disagree."

Such acts of resolution are made necessary, according to the present theory, by the situational demands of co-orientation on the one hand and by the psychological strain toward symmetry on the other. But symmetry is only a facilitating condition for co-orientation, not a necessary one. While (as maintained in the preceding propositions) the probabilities of symmetry vary, *ceteris paribus*, with demand for co-orientation, the theory does not demand that a symmetry-directed AtoBreX occur in every instance of strong demand for co-orientation. On the contrary, the theory demands that it occur only if, as, and when co-orientation is facilitated thereby. We must therefore inquire more closely into the nature of the forces toward co-orientation as related to possible forces against symmetry.

One kind of situational variable has to do with the nature of the forces which result in association between A and B. Of particular importance are constrained (enforced) vs. voluntary association, and association based upon broad as contrasted with narrow common interests. The range of X's with regard to which there is demand for co-orientation is presumably influenced by such forces. The relevant generalization seems to be as follows: *The less the attraction between A and B, the more nearly strain toward symmetry is limited to those particular X's co-orientation toward which is required by the conditions of association.* This would mean, for example, that as attraction between two spouses decreases, strain toward symmetry would increasingly narrow to such X's as are required by personal comfort and conformity with external propriety; similarly, the range of X's with regard to which there is strain toward symmetry is greater for two friendly than for two hostile members of a chess club.

The problem of constraint has already been noted. In some of the stud-

ies cited above it was assumed that frequency of communication varies with attraction, but this is not necessarily true under conditions of forced association. Two recent theoretical treatises deal with this problem.

Homans, one of whose group variables is "frequency of interaction" (though not communication, specifically), includes the following among his other propositions: "If the frequency of interaction between two or more persons increases, the degree of their liking for one another will increase, and vice versa"; and "The more frequently persons interact with one another, the more alike in some respects both their activities and their sentiments tend to become" (7, p. 120). (The latter proposition, which closely resembles the one here under consideration, apparently takes a much less important place in Homans' system than the former.) Almost immediately, however, the latter proposition is qualified by the statement, "It is only when people interact as social equals and their jobs are not sharply differentiated that our hypothesis comes fully into its own." In nearly every chapter, moreover, Homans (whose propositions are drawn *post hoc* from various community, industrial, and ethnological studies) points to the limitations which are imposed by constraining forces—particularly those of rank and hierarchy—upon the relations among attraction, similarity of attitude, and communication.

Blake manages to incorporate these considerations in a more rigorous proposition. Noting that hostility cannot be considered as the simple psychological opposite of positive attraction, he proposes to substitute a curvilinear for Homans' linear hypothesis: ". . . when pressures operate to keep members of a group together, the stresses that drive toward interaction will be stronger in *both* positive and negative feeling states than in neutral ones" (2). This proposition seems consistent with the present argument to the effect that demands for co-orientation are likely to vary with the nature and degree of constraints upon association; hence communicative acts, together with their consequences, will also vary with such constraints.

Another situational variable deals with the fact that, under conditions of prescribed role differentiation, symmetry may take the form of "complementarity" (cf. 11) rather than sameness. For example, both a man and his small son may (following a certain amount of communication of a certain nature) subscribe to the *same norms* which prescribe *differentiated behavior* for man and boy with respect to a whiskey and soda. If the father drinks in the son's presence, there are demands upon both of them for co-orientation; but there is strain toward symmetry only with respect to "the code," and not with respect to personal orientation toward the whiskey and soda. The code becomes the X with regard to which there is strain toward symmetry. In more general terms, *under conditions of differentiation of A's and B's role prescriptions with regard to X, the greater the demand for co-orientation the greater the likelihood of strain toward symmetry with respect to the role system* (rather than with respect to X itself).

A third situational variable has to do with the possibility that symmetry may be threatening. Particularly under circumstances of shame, guilt, or fear of punishment there are apt to be strong forces against a symmetry-directed AtoBreX, even though—in fact, especially when—attitude toward X (the guilty act) and attraction toward B (a person from whom it is to be concealed) are strong. Under these conditions it is the demand for co-orientation which creates the problem; if A could utterly divorce X (his own act) from B, he would not

feel guilty. Forces toward symmetry, however, are opposed by counterforces. Demand for co-orientation induces strain toward symmetry, but does not necessarily lead to a symmetry-directed AtoBreX.

A theoretically analogous situation may result from the omnipresent fact of multiple membership groups. That is, strains toward symmetry with B_1 in regard to X may be outweighed by strains toward symmetry with B_2, whose orientations toward X are viewed as contradictory with those of B_1. This is often the case when, for example, two good friends "agree to disagree" about something of importance to both. Thus in one study (14) it was found that those members least influenced by reported information concerning their own group norms were those most attracted to groups whose norms were perceived as highly divergent from those of the group in question.

Communicative acts, like others, are thus subject to inhibition. Such "resolutions" as "agreement to disagree," however, represent relatively stressful states of equilibrium. It is therefore to be expected, in ways analogous to those noted by Lewin in his discussion of the quasi-stationary equilibrium, that A-B-X systems characterized by such stress will be particularly susceptible to change. Such change need not necessarily occur in the particular region of the system characterized by maximal strain.

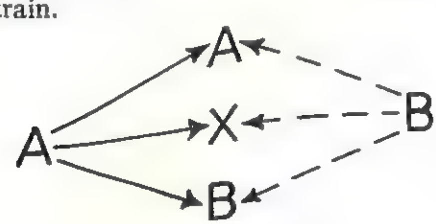


FIG. 2. Schematic illustration of A's phenomenal A-B-X system

The dynamics of such a system are by no means limited to those of strains toward symmetry, but must include changes resulting from acceptance of existing asymmetry. The possible range of dynamic changes is illustrated in Fig. 2. (In this figure, the A and B at either side represent persons as communicators; the A and B in the center represent the same persons as objects of co-orientation. The broken lines represent A's judgments of B's orientations.) Given perceived asymmetry with regard to X, and demand for co-orientation toward B and X, the possibilities for A are such that he can:

1. achieve, or attempt to achieve, symmetry with regard to X
   a. by influencing B toward own orientation,
   b. by changing own orientation toward B's,
   c. by cognitively distorting B's orientation;
2. introduce changes in other parts of the system
   a. modify his attraction toward B,
   b. modify his judgment of own attraction for B,
   c. modify evaluation of (attraction toward) himself (A),
   d. modify his judgment of B's evaluation of himself (B);
3. tolerate the asymmetry, without change.

As suggested by this listing of possible "solutions," the perception of asymmetry, under conditions of demand for co-orientation, confronts A with a problem which he can attempt to solve behaviorally (i.e., by communicative acts) and/or cognitively (i.e., by changing either his own orientations or his perception of B's orientations). Whatever his chosen "solution," it has some effect upon A's phenomenal A-B-X system—either to reinforce it or to modify it. As a result of repeatedly facing and "solving" problems of co-orientation with regard to a given B and a given X,

a relatively stable equilibrium is established. If A is free either to continue or not to continue his association with B, one or the other of two eventual outcomes is likely: (*a*) he achieves an equilibrium characterized by relatively great attraction toward B and by relatively high perceived symmetry, and the association is continued; or (*b*) he achieves an equilibrium characterized by relatively little attraction toward B and by relatively low perceived symmetry, and the association is discontinued. This "either-or" assumption under conditions of low constraint presupposes a circular relationship between attraction and the perception of symmetry. The present theory demands this assumption of circularity, and empirical evidence (under conditions of relative freedom from constraint) seems to support it.

Under conditions of little or no freedom to discontinue association, no such circularity is assumed. The conditions which dictate continued association also dictate the requirements for co-orientation, which are independent of attraction. The empirical data suggest that the degree to which attraction is independent of symmetry varies with the degree of *perceived* (rather than the degree of objectively observed) constraint.

Group Properties

It follows from the preceding assumptions and propositions that there should be predictable relationships between certain properties of any group and variables having to do with communicative behavior within that group. A group's structural properties, for example, viewed as independent variables, may create problems and may provide solutions to other problems of communication. Viewed the other way around, many properties of a group are outcomes of its communicative practices. Evidence from many sources points to distinctive properties of groups which are precisely those which the foregoing considerations would lead us to expect, either as conditions for or as consequences of a given kind and frequency of communicative acts.

Three kinds of properties are briefly noted. Each of them is hypothetically related (either as dependent or as independent variable) to the probabilities of the occurrence of a given kind of communicative act.

1. *Homogeneity of orientation* toward certain objects. All descriptive accounts of interacting groups note this property, in one way or another and by one label or another. As applied to behavior, it does not necessarily refer to similarity of action on the part of all group members, but only of demand or expectation; e.g., all expect each to take his own differentiated role. In order to account for the observed facts it is necessary to make the assumptions (not previously made in this paper) that information may be transmitted in nonverbal ways, and with or without intention to do so—e.g., a person's behavior with regard to a given object informs observers about his orientation to it.

If communication is thus broadly defined, then the degrees of homogeneity of orientation of a given group with respect to specified objects are presumably related to communication variables with respect to those objects. It is not hypothesized that homogeneity is an invariable function of any single index of communication (frequency, for example), but rather that it varies in accordance with the dynamics of A-B-X systems. While there are often extra-group determinants of homogeneity of orientation, it seems reasonable to view this very important group property as an outcome of the conditions and consequences of communicative acts.

2. *Homogeneity of perceived consensus* (i.e., homogeneity of judgments

of homogeneity of orientation). This property, though not often specifically mentioned in the literature on groups, is usually implicitly assumed. Most communication presupposes a considerable degree of perceived as well as objective homogeneity of orientation. The very fact of using language or gesture presupposes the assumption of consensus among communicants as to the information transmitted by the use of symbols.

Homogeneity of orientation and of perceived consensus do not, in spite of implicit assumptions to the contrary, have an invariant relationship; judgments of homogeneity may be of any degree of accuracy. If, as in the village reported by Schanck (13), each of many dissenters from a supposed norm believes himself the only dissenter, this state of pluralistic ignorance is an important group property, and is plausibly described by the author as an outcome of certain practices of communication. Any degree of homogeneity of perceived consensus, representing any degree of accuracy, is hypothetically an outcome of previous communicative acts and a determinant of future ones.

3. *Attraction among members.* Relationships of positive attraction of some degree invariably characterize continuing groups under conditions of minimal constraint, and are commonly found even under conditions of considerable constraint. This is so commonly the case that Homans (7) ventures the almost unqualified hypothesis that "liking" increases with frequency of interaction, and vice versa. Viewed in the light of the hypothetical dynamics of A-B-X systems, Homans' proposition would be amended to the effect that interpersonal attraction varies with the degree to which the demands of co-orientation are met by communicative acts.

These are not, of course, the only group properties of significance, nor are these properties outcomes exclusively of intragroup communication. (Some properties of almost any group, particularly at early stages of its history, derive largely from individual characteristics which its members bring to it.) It appears to be the case, nevertheless, that the hypothetical conditions and consequences of communicative acts are not limited to groups of two, and that some of the important properties of observed groups are consistent with the hypothetical dynamics of A-B-X systems.

SUMMARY

Communicative acts, like other molar behaviors, may be viewed as outcomes of changes in organism-environment relationships, actual and/or anticipated. Communicative acts are distinctive in that they may be aroused by and may result in changes anywhere within the system of relations between two or more communicators and the objects of their communication. It seems likely that the dynamics of such a system are such that from an adequate understanding of its properties at a given moment there can be predicted both the likelihood of occurrence of a given act of communication and the nature of changes in those properties which will result from that act.

Some of the most significant of group properties are those which, hypothetically, vary with intragroup communicative acts. It should therefore be rewarding to discover whether support for the present hypotheses, as apparently provided by the scattered evidence now available, can be confirmed in more systematic ways. If so, there are promising possibilities of investigating the phenomena of social interaction by viewing them as events within communication systems.

REFERENCES

1. BACK, K. The exertion of influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.
2. BLAKE, R. R. The interaction-feeling hypothesis applied to psychotherapy groups. *Sociometry*, in press.
3. FESTINGER, L. Informal social communication. In L. Festinger, K. Back, S. Schachter, H. H. Kelley, and J. Thibaut, *Theory and experiment in social communication*. Ann Arbor: Institute for Social Research, Univer. of Michigan, 1950.
4. FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups*. New York: Harper, 1950.
5. FESTINGER, L., & THIBAUT, J. Interpersonal communications in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92–99.
6. HEIDER, F. Attitudes and cognitive organization. *J. Psychol.*, 1946, **21**, 107–112.
7. HOMANS, G. C. *The human group*. New York: Harcourt Brace, 1950.
8. HOROWITZ, M. W., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1951, **4**, 57–76.
9. MILLER, G. A. *Language and communication*. New York: McGraw-Hill, 1951.
10. NEWCOMB, T. M. *Personality and social change*. New York: Dryden, 1943.
11. PARSONS, T., & SHILS, E. A. (Eds.) *Toward a general theory of action*. Cambridge: Harvard Univer. Press, 1951.
12. SCHACHTER, S. Deviation, rejection and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.
13. SCHANCK, R. L. A study of a community and its groups and institutions conceived of as behaviors of individuals. *Psychol. Monogr.*, 1932, **43**, No. 2 (Whole No. 195).
14. WHITE, M. S. Attitude change as related to perceived group consensus. Unpublished doctoral dissertation, Univer. of Michigan, 1953.

[MS. received December 15, 1952]

Psychological Review
Vol. 60, No. 6, 1953

ON SOME CORRUPTIONS OF THE DOCTRINE OF HOMEOSTASIS

J. R. MAZE

University of Sydney

Cannon's doctrine of homeostasis (2) simply asserts that many enduring and apparently static physiological states are in fact the product of a constant opposition of forces—an interchange of substances, a constant adding-to and taking-away—and that with some of these, when there is a departure from the balance then that departure sets in motion processes which (under specifiable conditions) restore the balance. (Although the action of many of these physiological mechanisms was first indicated by Cannon, the discovery of similar "constant inconstancies" had been made many times before, e.g., by Heraclitus, who contended that anything at all was made up of such steady states of flux.)

But in many recent treatments of homeostasis all that is retained is just the steadiness, the tendency to restoration. The constituent opposing processes are lost sight of, and so are the compensating mechanisms, and the suggestion is conveyed more or less explicitly that the mere departure from the condition in question is sufficient to restore that condition—as one might unreflectingly think that the mere deviation of a pendulum from the vertical (as a result of some external force) is sufficient to start it on its way back to the vertical (almost as though there were something "illogical" or "unnatural" about any other position). When we consider the number and variety of places from which any given state—say, a certain temperature or a certain concentration of carbon dioxide—is absent, then it becomes plain that mere absence alone is never sufficient eventually to produce that condition. We must consider what is present in, as well as absent from, the place we are talking about, what structures and processes are there to act on each other and so produce the effect being investigated. Just the same point can be made about doctrines which use the principle of "closure" to explain the emergence of certain perceptions, or motor actions, or neural patterns.

Many instances of these opposing processes and their attendant mechanisms are discussed by Cannon. To take only one example, one relatively constant physiological state is the concentration of carbon dioxide in the blood stream. Carbon dioxide is continually being produced, primarily by the combustion of lactic acid in the working muscles, and continually passing out of the blood through the lungs. When, for any of a number of reasons, one of them—the production, say—outstrips the other, so that the concentration of carbon dioxide rises, this increase has among its many effects that of exciting certain nerve endings in the carotid sinus which produce direct, reflex stimulation of the musculature to produce more vigorous breathing, by which the rate of excretion of carbon dioxide in the lungs is increased, so that, under specifiable conditions, the carbon dioxide concentration falls. It is misleading (although perhaps in fact correct) to say, as is often done, that it falls until it passes back below the "danger point," whereupon the respiration returns to normal. What happens, of course, is that it falls below the threshold at which those nerve endings are excited, so that the extra impulses are no longer delivered to the respiratory muscles, which

naturally then return to their former level of work.

Now, to give such a balancing the name "homeostasis" seems unobjectionable, but the mere existence of the word provides another opportunity for us to give up the tiresome search for those mechanisms which actually produce the effect in question, the constancy, and simply to attribute its production to the activity of a vaguely conceived and therefore very accommodating force named "homeostasis." Nobody would, of course, perpetrate this fallacy in such a crass fashion; it is hard to convict people of it because its very crudity leads them unconsciously to shy away from it and disguise it in ambiguous phrases. Yet as Lashley (5) pointed out, theories of motivation are riddled with such occult powers.

Even Fletcher (4), who professedly wants to exhibit homeostasis as an explanatory principle, says nothing which is unequivocally open to this objection, although one might contend that if he is not speaking of homeostasis as a force which produces constancies, then he is really not saying much at all. For example, concerning the retinal changes in color vision, as sketched by the Hering theory, he says, "We, of course, do not know much about the original organic status, nor precisely how it is disturbed, nor yet concerning the process of recuperation. Yet we cannot but assume the existence of such a status, the disturbance of the status, and a process which brings about its restoration" (4, p. 84). Since he offers this as an example of the use that "can be made of the principle of homeostasis in dealing with some well-known, but as yet not well-explained mental phenomena," it would appear that the "process" which restores the status is just homeostasis itself. But of course homeostasis is simply the *fact* of restoration, the word is only another name for the restoring, and so it cannot be appealed to as something which explains that restoration.

This point appears again when Fletcher refers to the J. J. B. Morgan experiments on the effect of distraction. "The now familiar fact that such distractions do not always decrease, but may even increase performance is taken to indicate a tendency to maintain a certain work level" (4, p. 85). Now, the fact that the subject tends to maintain a certain work level is not what is indicated but just what is *observed*—that is, that he usually *does* maintain it. This alleged explanation then merely restates the fact to be explained.

Homeostasis, then, is to be seen not as the *cause* of "steady states," but as the steadiness itself, as the *effect* of various specific qualitative processes in the organism. Nor can references just to the activity of "the organism" replace the discovery of those specific processes or mechanisms. A proper discussion of this issue is beyond the scope of this essay, but at least one can contend that the opposition between the organismic (or holistic or Gestalt) view and the atomistic view is not, as Wheeler (10) claims, exhaustive of the possibilities. We are not required to see either the whole as primary and prior to the parts or the parts as primary and prior to the whole. The solution is that both whole and parts must be recognized as having distinct, even if not separable, existence. The organismic approach is of course not to be discarded completely; in brief, we must recognize that there *are* wholes having their own characteristics. But an extreme organismic view, holding that no statement which treats the parts as distinct from each other and from the whole can be quite true, is faced with the problem of showing how there can be said to be a whole of parts (in the plural) at all. If the parts are not distinct from each other and from

the whole then they are identical with it, that is, it cannot actually be said to *have* parts, that is, it is indistinguishable from an atom, and Wheeler's disjuncts are not only not exhaustive but not even exclusive.

Statements, then, such as Stagner's that "when a given tissue constancy is first disturbed, the organism mobilizes energy for action which ceases when the equilibrium is restored" (7, p. 5), can be quite misleading, even though they may be intended only as preliminary and liable to further elaboration. This elaboration, if it is ever offered, must show that it is not that the organism mobilizes energy for action, but only that certain parts of it are caused to act in specific ways, either by circumstances external to the organism or by the action on them of other parts of it. There is a special case in which this objection seems not to apply—namely, when we are speaking of psychological actions involving intentions and using knowledge gained in past experience. But the notion that establishing homeostasis is something that "the organism" does is carried over into the field of physiology, and although once again it may itself seem unobjectionable if a trifle vague, it readily leads to the suggestion that organisms possess or develop various mechanisms *in order* to achieve homeostasis. Putting it another way, the notion arises that organisms possess a special kind of causality, namely, a teleological causality, so that the part-processes within an organism are determined not by the *nature* of those parts (since they have no distinct nature), but by the effect which *is to be produced* by them in the organism as a whole.

Anything more than a suggestion of this is rarely found, but nevertheless we do find these suggestions even in Cannon's use of such phrases as "the means employed by the more highly evolved animals for preserving uniform and stable their internal economy" (2, p. 24); and in Dempsey's ". . . in order to free themselves from their marine habitat, they had to devise mechanisms for maintaining the fluid concentrations of the body" (3, p. 230). Again, when Fletcher, speaking of the capacity of the physical organism "to combat invading foreign bodies and pathogenic organisms," says that "the adaptability of this capacity to deal with novel conditions is suggestive of a form of intelligence" (4, p. 80), he is thinking of the physiological events as means adopted for a certain end, rather than as events whose effects happen to have survival value for the organism in question, that is, as long as it happens to remain within a certain kind of environment.

How would we deal, on this view, with the evidence Fletcher refers to that "the introduction of foreign substances into the blood stream of certain animals provoked the formation of antibodies specifically qualified to destroy these substances, even though, presumably, the animals had never been exposed to such substances in the course of their evolution" (4, p. 81)? One point would be that this absence of previous exposure is quite irrelevant; what is relevant is whether they previously possessed any, no matter how few, of the organisms which are now called antibodies. Another important point is that this capacity to form antibodies is highly selective for different foreign substances; otherwise, we could not understand how these animals could ever be poisoned, or die of a contagious disease. What determines this difference, that some foreign substances are destroyed and others not? In the former case there must already be present in the animal something, presumably some kind of organism, for whose propagation the presence of the foreign sub-

stance provides favorable conditions—and how often do we find that "providing favorable conditions" means being consumed by, that is literally being *food* for, the thing in question? There is nothing surprising in the presence of a certain crop or species of game resulting in the increase, in that area, of the human race—organisms "specifically qualified" to destroy that crop or game.

There is nothing in homeostasis which requires a special kind of organic causality, nothing which cannot be rendered in statements of antecedent conditions and consequent effects, nothing, in fact, peculiar to organisms at all. The thermostat is often used as a physical analogy. An even more primitive but most effective homeostatic mechanism is the outrigger of a canoe. As the canoe begins to tip over to one side, because of the action of the waves, for example, the float of the outrigger begins to rise in the water, its weight increases, and this swings the canoe back. If it goes too far, the float is pressed down into the water, the upward pressure on it is increased, and again there is a return toward the even keel. These mechanisms, of course, happen to be arranged by men, but no foreseeing of ends is necessary for homeostasis; any floating log with a projecting branch exhibits just the same behavior as that indicated in the canoe with its outrigger.

One point arising in each of these examples makes it plain that homeostasis is not just one process, and so that it cannot be dismissed just with a reference to "the organism." That is, there are always at least two mechanisms, two causal sequences, involved: if two forces A and B are opposing each other in a homeostatic situation, then one mechanism or process will be brought into play when A outstrips B and a different (although often intimately related) one when B outstrips A. These may be processes internal to A or B rather than something opposing them externally; e.g., it might simply be that A can only stretch out, as it were, to a certain extent and then becomes exhausted—a principle which seems to be clearly enough recognized in warfare, supposing A now to be an advancing army. This internality may give the impression that homeostasis *is* a force in itself, that there *is* something illogical about the departure from balance, but actually we can only understand this internal necessity by realizing that A (*whatever* it may be) is articulated, has distinct parts, and that the "internal necessity" is an external action of those parts on one another. In this instance we might have to take notice, e.g., of a lack of coordination between the general staff dictating the speed of advance and the supply corps trying to keep up the materials consumed by that advance.

Now, a further important point, and one which in my opinion renders otiose many "uses" of the principle of homeostasis, is that what is restored is not just "equilibrium," as in the phrase I quoted from Stagner, but some specific amount or concentration or intensity of a particular property. That is, what we have to recognize is not just equilibrium or constancy, but *kind*. Thus the outrigger keeps the canoe not just steady but right-side-up; the thermostat maintains not just any steady temperature but one within a specified range. If it were possible for the many relevant mechanisms to stabilize the pH of the blood at a point outside the range 6.8 to 7.8, then death would speedily ensue. Plainly in this case, if we were interested in the organism's future, it would be of crucial importance to know not merely that the pH was constant, but at what point it was constant.

Stagner, in the article in question

(7), and in his recent text with Karwoski (8), seems to have become the chief exponent of homeostasis as a foundation for psychological theorizing, and, I would contend, the leader in concentrating on constancy and neglecting to observe what the constant states are. In proposing homeostasis not as an explanatory principle, but just as a "unifying concept," he adopts the position that all goal-directed behavior originates in the endeavor to maintain tissue constancies. After some experience the organism anticipates disturbance, and "perceives environmental objects as potential sources of equilibrium-restoration" (7, p. 5). One of the means-activities directed towards this end is the adoption of perceptual constancies. "Objects, as stimuli affecting the distance receptors, are protean in size, shape, and color. Under such conditions adjustment is most difficult. The organism therefore *learns* to perceive identical objects as possessing these constant attributes" (7, p. 7).

Concerning the size, shape, and color constancies, it seems plain (if we adopt a realist position) that the young organism is developing its ability to recognize, under varying conditions, properties which the object actually does possess throughout the period in question, despite what naively appear to be changes in them. (One might contend that unless the organism had some *native* ability to recognize these constancies there would be no possibility of its learning to perceive them. If it were able to bring together the various appearances belonging to identical objects, as Stagner suggests, then it must *already* have been able to recognize each of these allegedly protean objects. Not only would there then be no need for the organism to develop those specific object constancies of shape, size, and color, but also no reason why it should not assume that the objects simply did continually change their shape, size, and color. Where would it get any other hypothesis?)

However, when Stagner contends that the demand for perceptual constancies is transferred to the perception of people, he apparently means not only that their *constant* features are perceived as being constant, but that features which they actually possess only intermittently are also to be perceived as being constant. Thus he says: "If I observe Mr. Smith behaving in a weak, futile, ineffectual manner today, I shall be predisposed to observe those same characteristics in his actions tomorrow (the so-called 'halo' effect). This phenomenon is likewise adaptive and homeostatic in character. . . . Since the constancy hypothesis proves useful in dealing with inanimate objects, it tends to be transferred to dealing with people" (7, p. 8).

Note that the perception of objects as having constant shapes, sizes, and colors has now become the adoption of an all-inclusive "constancy hypothesis," referring to all their properties whatever. This is one of the many ways in which Stagner ignores the *content* of constancy. As far as those original object constancies are concerned, it is not having a *constant* perception but a *correct* one that is important, but Stagner is intent on constancy whether or no.

This "constancy hypothesis" then is held to be useful in dealing with people. Thus, he says: ". . . successful adaptation of the child to demands of his parents is unquestionably facilitated if he behaves on a constancy hypothesis with respect to them. When reality constantly frustrates this tendency (very inconsistent real behavior by a parent), maladjustment seems invariably to result" (7, p. 8). That is, in contradiction to his preceding sentence, the "constancy hypothesis" would *not* facilitate adaptation. Of course, the point is that Stagner fails to distinguish between in-

consistency and variability. Once we see that there is no inconsistency in the sense of illogicality, we realize that the parent's behavior, no matter how much it may change from day to day, always exhibits regularities. Reality, then, cannot frustrate the tendency to "perceptual constancy" in the sense that it could make the discovery of regularities—whether these be the possession of permanent predicates or the lawful recurrence of intermittent ones—intrinsically impossible. Of course, the parent's behavior may sometimes vary in a way which the child is not yet able to understand, but I suggest that it is not such failure to understand that leads to "maladjustment." The psychoanalytic view that such variability really arises from ambivalent feelings towards the child, and that it is the child's recognition of the underlying hostility which leads to "maladjustment" or anxiety, seems to me to be much more plausible and fruitful. Be that as it may, the question is always not whether something is constant, but what feature it is that it constantly exhibits.

Stagner does appear to recognize the qualitative content of constancy when he speaks of "constancies," i.e., in the plural: ". . . the organism is constantly faced with the necessity for maintaining a variety of constant states" (7, p. 10). But one might ask here why bother to use the term "constant" at all? *Any* state is constant for the time it endures, and if "constant" is here supposed to mean anything more than that, if the states really *were* "constant," then there would be no need for "the organism" to maintain them at all. However, the plurality and the qualitative nature of these constancies soon disappear when Stagner finds a common denominator for them, a general equilibrium. "The facts indicate," he says (7, p. 11), "that the epithet 'coward' may be more disturbing to equilibrium than the physical danger of facing a wild animal." But what are the forces that are *in* equilibrium? What is the resultant condition of their equilibration? One must recognize an inexhaustible (and irreducible) variety of equilibria in any organism, and while one might conceivably compare the magnitude of a disturbance in one with that of a disturbance in another, still these disturbances are distinct from one another; they are disturbances of qualitatively diverse states, and qualitatively diverse conditions result from them.

The plurality of opposing forces in the organism is finally obscured, paradoxically enough, when Stagner comes to deal with the question of choice. He says that this need for maintaining a variety of constant states faces the organism with "the task of determining priorities. . . . It should be clear . . . that the organism in some way evolves standards of value, in terms of which choice is made as to the particular constancy which gets priority" (7, p. 11).

Admittedly, it is especially easy to evade searching for the mechanisms involved when we are examining psychological or mental homeostasis, since our knowledge of even the most general nature of mental structures is negligible.

Some system of tensions, however, seems to be the general structure of mind, where the tensions are of a variety of kinds, or better, where there is a variety of structures, each becoming tense. There may be outlets (activities) which release in varying degrees the tension of more than one structure, and any one body of tension will have a variety of outlets, so that when one is blocked, e.g., by "conscience," then the tension rises until it flows through another (or perhaps by accumulation becomes strong enough to burst through the first barrier). But the point is that, because of their qualitative diversity,

the releasing of a tension is not a diminution of some undifferentiated body of energy of which all activities are a manifestation, but is always the releasing of one *specific* tension, leaving others quite untouched.

Now, what makes them a "system" is that, usually because of social pressures, the releasing of one tension can throw up obstructions which prevent the release of another. The intricacy of any given moral code is an indication of the intricacy of the web of contingencies that can grow up between various impulses. (The word "impulse" is used to indicate a structure having tension. Even though a structure may never be devoid of tension, still those two features of an "impulse" are distinguishable.) When we say we are in a state of indecision (as Stagner puts it, "the organism is faced with the task of determining priorities"), what is really happening is that two impulses or groups of impulses of approximately equal strength have come into conflict, and each is enlisting still further impulses on its side. The main means of effecting such recruitment is presumably the gaining of knowledge, or rather the acquiring of beliefs. One might put it that a given impulse *X* comes to believe, or is, as it were, assured by impulse *Y*, that the "satisfaction" (release of tension) of *X* is facilitated by the satisfaction of *Y* and opposed by the satisfaction of *Y's* opponent, *Z*. "Choice," then, is just the fact of one group of impulses simply becoming stronger than the other and forcing its way into action past that other group's obstruction.

Some such view is essential to getting rid of the mystification of Reason or the Will, the "ghost," as Gilbert Ryle calls it, and the Organism, as Stagner calls it. One cannot explain why a man "makes up his mind" one way or the other by attributing the decision to some single executive, since precisely the same problem breaks out afresh when we ask why that executive went one way rather than the other. Once again, in explaining activity we have to look for the action of the parts of the active thing on one another, and in general we might say that anything which can throw its weight into and alter the course of a conflict of impulses thereby becomes simply another coordinate impulse.

Of these, Stagner actually recognizes only the physiological processes, the "tendency to tissue constancy." Where the person seems to have a conflict between psychological and physiological impulses he is really just making a kind of mistake. Thus, in dealing with "the artist who starves in a garret to express his ideas, the martyr who prefers physical pain to an abandonment of his beliefs," Stagner suggests that this depends on "the evolution of a new perceptual object, the self or ego, and the establishment of a perceptual constancy with regard to it. . . . The tendency towards perceptual constancy will operate to hold constant the self-image" (7, p. 12). But what we really want to know is what holds constant the *self*, why does the artist go on being an artist? Stagner's theory can at most only tell us why the artist goes on *thinking* of himself as an artist, i.e., because the tendency to perceptual constancy, which in the first place was only a means to the goal of tissue constancy, has gained some sort of autonomy, has gone astray, and no longer contributes to that original goal. This intellectualism, in which the having of a certain notion of oneself is taken as sufficient to lead one toward being that sort of person, crops up again and again throughout the article. Stagner does say later that the self-image becomes so influential because the maintenance of a given ego status is important for "the maintenance of all needed equilibria," and so the self-image

is somehow "reinforced by" different "inner tensions." But if this is so, then any tendency to perceptual constancy as such is quite superfluous; the specific tensions are sufficient. Furthermore, what is this ego of which there is an image, and what *has* the image? The ego is simply that system of "inner tensions" themselves, and the clinging to a self-image is simply these various tensions being aware of each other as allies or as opponents. One never has an image of oneself as a totality, which one then seeks to attain or maintain. It is always that certain specific features are approved, and others disapproved. And the question arises again, what approves and disapproves? Not "the ghost," not some transcendent ego, but merely various component impulses struggling for outlet, and both expressing and furthering that struggle with these feelings of approval and condemnation.

The play of impulses in the mind might well be likened to the struggle of political parties, religious institutions, industrial factions, and the like in a society—a field in which the guiding hand of homeostasis has also been discovered, e.g., by Cannon himself (2) and by Dempsey (3). It seems fairly plain that such measures as the regulation of production in order to dampen the cycle of boom and depression, the granting of award wages, the legal freedom of worship are not adopted by the community for the good of all, but are produced by the fact that each of the various social groups is fanatically pursuing its own interest, intriguing, forming and dissolving alliances, calling strikes and lockouts, and being forced into grudging and temporary compromises when it is unable to muster enough strength to win a clear victory. The rough stability of a society over a period of time is not the result of the society's striving to retain its identity, or even its "acquiring its own momentum," but is just a by-product of the many-sided clash of sectional interests. There is nothing inexorable about their always balancing each other: some struggles are won and lost; societies do change in various respects and keep on changing and sometimes collapse and disappear; and so the clashing impulses in individual men sometimes balance each other and sometimes do not; men do change as their lives progress, stop being artists, embrace religion and lose faith, abandon their neuroses and commit suicide.

The doctrine of homeostasis at best only points to the facts of opposition and cooperation without advancing knowledge of the impulses whose activities these are; at worst, it hinders that inquiry by ignoring those impulses and concentrating on their resultant (or even on their mere equilibrium), and by offering a pseudosolution of how the more or less stable resultants are maintained.

REFERENCES

1. Anderson, J. Mind as feeling. *Aust. J. Psychol. Phil.*, 1934, 12, 81–94.
2. Cannon, W. B. *The wisdom of the body.* London: Kegan Paul, 1932.
3. Dempsey, E. W. Homeostasis. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 209–235.
4. Fletcher, J. M. Homeostasis as an explanatory principle in psychology. *Psychol. Rev.*, 1942, 49, 80–87.
5. Lashley, K. S. Experimental analysis of instinctive behavior. *Psychol. Rev.*, 1938, 45, 445–471.
6. Morgan, J. J. B. The overcoming of distraction and other resistances. *Arch. Psychol.*, 1916, No. 35.
7. Stagner, R. Homeostasis as a unifying concept in personality theory. *Psychol. Rev.*, 1951, 58, 5–17.
8. Stagner, R., & Karwoski, T. F. *Psychology.* New York: McGraw-Hill, 1952.
9. Weber, C. O. Homeostasis and servomechanisms for what? *Psychol. Rev.*, 1949, 56, 234–239.
10. Wheeler, R. H. Organismic vs. mechanistic logic. *Psychol. Rev.*, 1935, 42, 335–353.
11. Young, P. T. *Motivation of behavior.* New York: Wiley, 1936.

[MS. received December 26, 1952]

Psychological Review
Vol. 60, No. 6, 1953

THE BRAIN ANALOGY: TRANSFER OF DIFFERENTIATION [1]

H. EDGAR COBURN

Registered Civil Engineer, San Diego, California

A previous paper (3) introduced the association tracts (AT's) and dealt with three examples which involved a single US. The present Brain Analogy (BA) paper is based on a phenomenon, transfer of differentiation, which appears to involve two separate US.

Fundamentally, the BA is a circuit arrangement which has a potential connection from every sensory receptor to every motor cell. The potential connections, known as delta cells, and the sensory receptors have the magnitudes of their various properties distributed throughout the populations of delta cells and receptors in an approximately normal manner. Stimuli are converted into pulses in both the sensory and motor cells. It is approximately correct to say that the coincidence of a sensory and motor pulse at the delta cell's connection to the motor cell is the event which converts the potential connection into a functional one; subsequent stimulation of the appropriate sensory receptor causes the attached delta cell to evoke a CR from the motor cell (1, postulates). Also, there is a property of the delta cells, called the disabling function, which generally limits the number of simultaneously-active functional connections to one on any motor cell. These functional or conditioned connections are occasionally redundant, but the first conditioned delta cell into action disables the others on the same motor cell in so far as the effect of reinforcement is concerned. As a result of this property, stable conditioned connections can exist only if the action of the delta cell is initiated prior to the action of the motor cell and is continued until the action of the two cells overlaps. Backward conditioning can have but a transient existence because there are always random S, some of which are in action prior to motor cell activity, that gain control of the motor cell and disable more latent delta cells. We see here part of the competition between delta cells which forms the basis of stimulus differentiation in an appropriate experimental program; the other factor is selective extinction. Continued or repeated action of a conditioned delta cell without stimulation of the receptors of the contiguous motor cell causes extinction—returns the delta cell to its original potential-connection status.

The distribution of the magnitudes of the properties of the sensory system is the responsible agent in the automatic creation of significant or rational functional connections between the sensory and motor systems. The delta cells that are actuated at the correct time are eligible to compete in the appropriation of the motor cells and capable of holding them against the encroachments of all other delta cells. The mechanism operates on a probability basis so that if a signal is followed by reinforcement the components of the signal actuate populations of delta cells in proportion to the magnitudes of the components, and the motor cells are appropriated in the same proportions; later application of the signal evokes the appropriate CR. If a component of the signal is applied separately it will be found that the evoked CR is roughly

[1] The author wishes to express his gratitude to Vera Jane Coburn for making the drawings and for assistance in preparing the manuscript.

proportional to the magnitude of the component and is the result of the activity of the particular motor cells appropriated by that component.

The AT's were introduced (3, first two sections) to cope with more complex phenomena than can be explained by simple sensory-motor (SM) conditioning. The AT's are merely additional branches of the sensory cells which lead to other sensory cell bodies where connections are effected by means of delta cells in a manner exactly the same as with connections to motor cells. As a result of these auxiliary connections it becomes possible to establish conditioned connections from one sensory function to another as in sensory preconditioning. With the AT hypothesis it was found desirable to limit generalized connections to the AT's, and hence all generalized CR's are effected from generalized sensory neurons through specific sensory neurons to motor neurons.

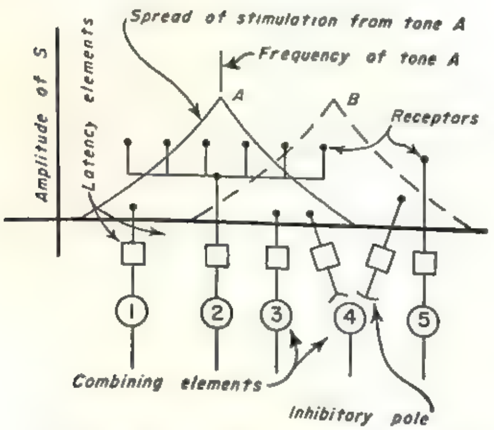


FIG. 1. The spread of auditory stimulation is possibly due, in part at least, to the fact that the basilar membrane deflects and stimulates receptors embedded in areas adjacent to the "point" of stimulation. This "generalization" is probably of insufficient degree to account for the facts without support from nonperipheral mechanisms. The intent in these figures is to simulate function rather than structure.

In this paper generalization is assumed sufficient to permit any overt S to act as a substitute for another. If we are differentiating two auditory S, say, it is probable that certain receptors will be actuated by each S unless they are low in strength and far apart in frequency. Suppose two tones, *A* and *B*, are of equal strength and less than an octave apart. Referring to Fig. 1 we see that, to the nervous system, the sphere of influence of each tone overlaps that of the other (4, p. 44).

The symbols used in the figures are the same as before (1). Briefly, the plus sign indicates a reinforced trial; minus, nonreinforced; zero, nonactuated; a parenthesis indicates an occasional trial; *S* and *L* equal short and long latencies, respectively.

TRANSFER OF DIFFERENTIATION

Pavlov's experiments on transfer of differentiation (5, p. 228) are concerned with two salivary functions. If these are entirely independent US, as implied by Pavlov, there is no ambiguity from that source in the following discussion. The experimental program consists in establishing tone *A*, say, as the CS for food and then differentiating tone *C* which has also become conditioned because of both generalization and closeness of physical relationship (4, p. 44). When the differentiation is established, *A* is given a series of trials in which it is followed by acid instead of food; this presumably causes the original salivary CR to become extinct and another qualitatively different salivary reflex becomes established. Upon testing again with *C* it is found that it does not elicit any response; the differentiation has been "transferred" to a new US.

Pavlov demonstrated that a new and finer differentiation could then be established by applying an intermediate tone, *B*, occasionally, and not reinforcing with

acid. This new differentiation was found to be transferable back to the original US, food.

Our S situation corresponds with Fig. 2 because overt *A*, *B*, and *C* are auditory S of less than an octave apart. None is ever actuated with another, *A* is always followed by food (or later, acid), and overt *B* and *C* are never overtly reinforced.

Figure 3 shows that S differentiation takes place partly at the AT level rather than at the SM level as indicated earlier (1, p. 173). Overt *A* actuates *1, 2, 7, 8, 9, 10,* and *13*. Neurons *2, 8,* and *10,* having long latency, are therefore objects of competition by neurons *1, 7, 9,* and *13*. The generalized group, *13*, is not a small population and no amount of application of overt *A* will cause differentiation since *1, 7, 9,* and *13* are components of a compound S as viewed by *2, 8,* and *10*. Yet differentiation is always available by rendering, say, *9* and *13* unstable—by applying occasionally overt tone *C* which stimulates *9* and *13* (among others) but not *1* or *7*; reinforcement by US_1 is omitted. Neurons *1* and *7*, which represent only the shorter latencies, become exclusive CS to neurons *2* and *8*, while *9* and the generalized connection, *13*, gradually become extinct at neurons *2* and *8* because of occasional nonreinforced (absence of *A*) trials.

In addition to *9* and *13*, *C* stimulates *5, 6, 10, 11,* and *12*. Only *10* concerns us now—the others will be disposed of later. Before any differentiation was effected, the group represented by *10* was appropriated by *1, 7, 9,* and *13* as indicated above. Differentiation, as a process, did not eliminate *9* and *13* from sharing control of *10* because the latter was also stimulated by *C* and therefore reinforced the AT connections that were actuated. Those portions of group *10* which were originally appropriated by

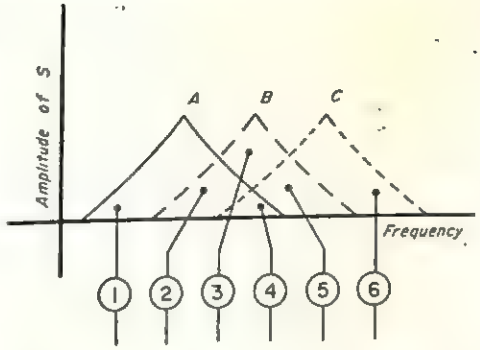


FIG. 2. Receptor properties, in effect, are partly a function of the magnitude and location in the sensory field of various stimuli and combinations of stimuli. For the particular program shown here, neuron *4* is generalized since *A*, *B*, or *C* will actuate it. Neurons *2* and *5*, on the other hand, are sensitive to only two of the stimuli. It is apparent that a reduction in magnitude of *A* and *C* will have the effect of changing *4* from a generalized to a specialized neuron. This dependence of properties on the experimental program is of considerable significance in altering the relative populations of delta cells available to the overt stimuli; the consequence is sometimes a marked difference in behavior.

1 and *7* were not disabled, since *1* and *7* are not actuated by *C*, and are therefore available for *additional* conditioned connections from any actuated, short latency, AT's: *5, 9, 11,* and *13*. Connections from *5* and *11* are supernumerary [2] with respect to applied *A* but *9* and *13* are redundant and probably most are extinctive since *1* and *7* were established competitively in step *1*. At the conclusion of differentiation the part of group *10* originally held by *1* and *7* is also held simultaneously by *5* and *11* in

[2] It is necessary to distinguish between redundant and supernumerary conditioning. Both are characterized by the presence of two or more conditioned connections to one sensory or motor cell body, but the conditioning is described as redundant only when two or more conditioned connections are in action simultaneously, otherwise it is supernumerary. In the latter case the connections are entirely without effect on one another in so far as reinforcement and the disabling function are concerned.

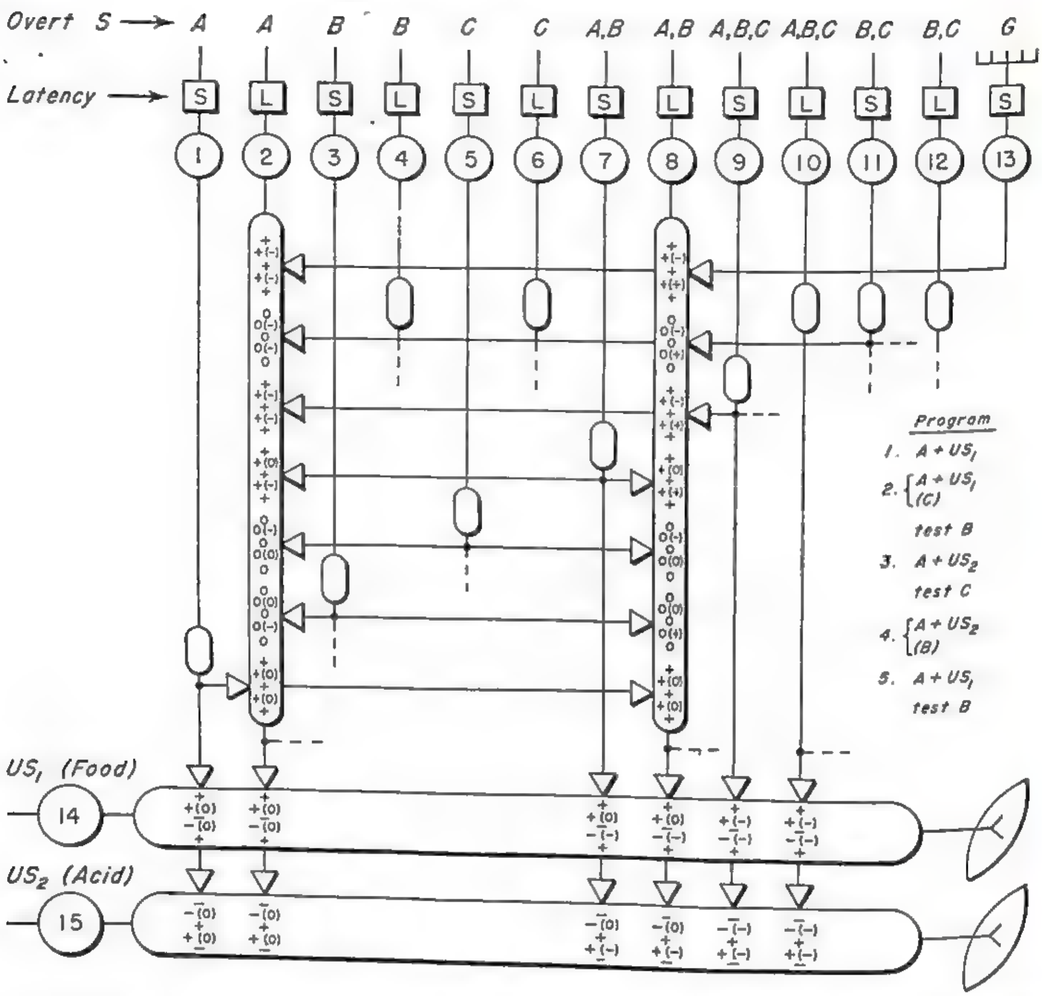


FIG. 3. This figure is necessarily a condensed version of the theory of transfer of S differentiation and can be interpreted properly only in conjunction with the argument in the text. The conditioning symbols should be read down the columns, each horizontal line corresponding to a numbered step in the program; no symbols are shown for the test trials interspersed between formal program steps.

four combinations. This is a failure of differentiation at the AT level. It will be seen, though, that differentiation drives to extinction the SM connections from *10* to *14* because *10* is actuated by any S and reinforcement by US_1 is not inevitable. SM connections from *1, 2, 7*, and *8*, not actuated when US_1 is absent, gradually gain exclusive control of *14*—no overt response follows application of overt *C*.

Thus far we have covered only the first two steps in the experimental program. Before proceeding further some digressions are necessary.

Allusion has been made before to the fact that AT conditioning is without significant effect unless the program is altered. A related property is the fact that sensory cells, as a group, must be excited directly rather than through AT connections in order to form relatively stable, numerically significant SM connections. The phenomenon does not seem to approach a limit, like habit saturation, and therefore a quantitative

approach is necessary. But this is impossible until suitable parameters have been established; for the present we place another burden on intuition.

Consider two sensory neurons, X and Y, and one motor neuron, Z. The program consists of trials of X and Y together, interspersed with trials of Y and Z together. During the former trials an AT connection develops from Y to X; during the latter trials an SM connection is made from Y to Z, or, if the AT connection is functional, chance will determine whether X or Y becomes conditioned. Actually, the AT delta cell and the Y-group SM delta cell do not have to be, and usually would not be, parts of the same neuron. Some other neuron, Y', for example, could gain control at the SM level. In any event, the AT and SM connections are subjected to an interspersed set of reinforced and nonreinforced trials: when the US is present the AT connection is not reinforced by X and when the US is absent the SM connections are driven toward extinction regardless of whether the sensory neuron considered is excited directly or through an AT connection. The circumstances of the assumed program are such that a group of X cells and a group of Y cells are in competition for control of Z, even though X cells are never excited directly when Z is actuated. The effective degree of competition is dependent on the nature of the program.

It will be observed that an X cell must be more latent (or be actuated later) than the Y cell which acts as a CS through the AT's. But this has no effect at the SM level because relative latencies are without influence, once a delta cell becomes conditioned, provided its action is initiated prior to that of the motor cell. The first *conditioned* delta cell into action disables all other delta cells on the same cell body (1, Postulate 13).

The circuit from Y through X to Z has less stability than that from Y to Z, directly, because of the two delta cells in series. The probability of encountering a short gamma phase cell is greater in the series path by a factor greater than one and less than two compared to the direct path.

The maximum *equality* of stability occurs when X and Y are usually actuated and the Y plus Z pair is seldom actuated. But this is the case when SM conditioning is nearly impossible to establish. The minimum equality of stability occurs when the above frequencies are reversed. In this case Y almost completely excludes X from Z. When the trials alternate, the Y connection to Z has roughly twice the stability of the X connection to Z because of the gamma phase distribution.

The typical differentiation program consists of reinforced trials interspersed with occasional nonreinforced trials. It is assumed that Pavlov's data (5, p. 228) correspond because no mention was made of any other procedure. With this assumption and the foregoing argument, it is seen that the neurons which are reinforced only when excited indirectly are also the ones that are excited infrequently; the resultant is a minimum equality of stability. In other words, neurons *4*, *6*, and *12*, representing overt B and C, cannot have significant access to *14* or *15* and therefore they are shown without input connections from the AT's in spite of the fact that such connections might be functional at times.

Neurons *1*, *3*, *5*, *7*, *9*, and *11* are shown without AT connections to their cell bodies because short latencies prevent stable conditioning from any other neurons in Fig. 3.

Returning again to the first step in the program, neurons *1*, *7*, *9*, and *13* are seen as components of a compound S as viewed by neuron *2*. They gain

joint control of *2* which, in the second step, becomes restricted to *1* and *7*. In the third step the latter two disable *9* and *13*, preventing renewed access. The fourth step drives the connection from *7* to extinction leaving *1* in sole possession of *2*.

Neuron *8* has a more complex history. The first three steps exactly parallel neuron *2* but the fourth step does not drive *7* to extinction because *8* is actuated by *B*. Neuron *8* then represents two portions. The first consists of those appropriated by *7* and continued in power by the program; the second consists of those appropriated by *1* which, not being actuated, cannot disable AT connections and therefore *3, 7, 9, 11,* and *13* gain access to *8*. The resulting conditioning from *7, 9,* and *13* is redundant in the fifth step but repeated application of *A*, in accordance with the program, eliminates the redundancy. The second portion of *8* is thereby consigned to neurons *1, 7, 9,* and *13* in proportion to their populations. Although the redundant connections are eliminated, the supernumerary ones are not; these are *3* and *11* which are silent partners of *1* in connections to *8*. It is easily seen that all or part of *8* can be actuated by overt *A, B,* or *C* after the fifth step in the program. With so little known concerning the mechanism that limits S differentiation, there is no recourse but to continue to appeal to the fact that differentiation *does* have limitations and it is not inconsistent to point to the relative populations as pertinent to the problem: when *A, B,* and *C* are too close in frequency the populations of *7, 8, 9,* and *10* are too large for effective transfer of differentiation.

Neuron *10* is actuated on every trial without exception. The explanation of of the resulting behavior of the AT conditioning is somewhat tedious and of little interest. It suffices to note that since *10* invariably responds it is necessarily driven to extinction in its connections to *14* and *15* if the latter two are sometimes not actuated and if any other sensory cell can acquire stable connections. These conditions are met in steps two and four, only. In the fifth step *10* aligns itself with *8* in opposing transfer of differentiation but it has a still smaller population.

The SM conditioning is relatively simple. In the first step neurons *1, 2, 7, 8, 9,* and *10* are components of a compound S as viewed by *14*. The second step eliminates the connections from *9* and *10*, transferring the unappropriated motor cells to *1, 2, 7,* and *8*. A test application of *C* reveals the differentiation but a test of *B* actuates *7* and *8*, and some of *2* through an AT connection from *7*, resulting in overt response of *14*. The third step drives all components to extinction on *14* and gives *1, 2, 7, 8, 9,* and *10* access to *15*. Since neurons *9* and *10* are not denied access to *15* it appears that transfer of differentiation has failed. And so it has, if *9* and *10* are large components. The magnitude of these components is roughly inversely proportional to the frequency difference between *A* and *C* with the result that the process approaches a limit as *C* approaches some value near *A* (in separate experiments). But while the SM differentiation was lost in the transfer, the AT differentiation remained untouched. Those outlets that *C* might have, by means of AT conditioning, through connections from *9* and *13* to *2* and *8* were driven to extinction before the transfer and subsequently were denied access by means of the disabling functions in *2* and *8*. As long as *9* and *10* are relatively small components transfer of differentiation is successful. As the frequency difference between overt *A* and *C* is reduced, in separate experiments, the populations of *9* and *10* increase and the ability to discriminate decreases. Finally, the

animal exhibits "defense" reactions and it is conjectured that these interfere with a smooth transition.

The fourth step in the program drives to extinction the SM conditioning from *7*, *8*, *9*, and *10* to *15* and the AT connections from overt *B* to *2* follow the same course. The fifth step breaks all connections to *15* and re-establishes connections from *1*, *2*, *7*, *8*, *9*, and *10* to *14*. But *1*, disabling all other AT connections to *2*, prevents a test application of overt *B* or *C* from causing a response through *2*. The direct connections from *7*, *8*, *9*, and *10* to *14*, as indicated earlier, oppose transfer of differentiation. If a considerable part of the response to tests is through AT connections to *2*, and these connections are broken, the response should decline perceptibly.

Since *9*, *10*, and *13* are in effect all generalized neurons, and since *9* and *10* have all the properties of *13* plus some additional properties, *13* seems superfluous. But this is decidedly not the case. The relatively large population of *13* has a significant function: it permits effective generalization through AT connections without at the same time preventing transfer of differentiation.

Transfer of differentiation is possible because of the existence of a relatively large population of generalized neurons which have no direct connection to any motor cell.

DISCUSSION

Several difficulties afflict the AT hypothesis as delineated in this and the preceding paper (3); three are mentioned here. First, it is apparent that if AT connections are made from short latency *A* neurons to longer latency *A* neurons during "primary" conditioning, *B* neurons cannot then gain access to any *A* neurons excepting those of short enough latency to avoid capture by other *A* neurons. While external access to *A* is thus severely limited, it may not be serious because of the connections internal to the *A* group.

Second, the AT's tend to reduce the latency of the overt response because a signal can avoid the delay of the long latency elements by a traverse through the short latency neurons and their AT's. The situation here is similar to the general case of a delayed CR. The reader can easily determine for himself that monopolar neurons are not adequate for delayed reflexes because all neurons being actuated for the same length of time regardless of latency, and long enough to overlap the US, are also subject to decrement for equal intervals. Those neurons possessing a long enough gamma phase gain control of the motor cells, but there is no differential action where latency is concerned and, therefore, the CR after training is actuated gradually and without any definite delay period. A nonresponsive delay period is assured by the simple expedient of excitatory bipolars which have two latencies individually subject to the usual distribution function. Those with a short overlap of paired latencies which is appropriately timed for the particular delay program are actuated briefly and only after considerable delay, hence they do not deplete the gamma phase and do not excite the response too soon; they are differentially favored on a latency basis in the struggle for control of the motor cells. It seems possible that the same mechanism may be applied to the AT's to prevent premature response.

Third, there are transient phenomena with which to cope. In Fig. 3, neuron *14*, it is shown symbolically that the connections from *1*, *2*, *7*, and *8* are occasionally not actuated on the second step of the program. In this case, and similarly for Fig. 4 of the preceding paper (3), the nonactuated stage is not reached until after the AT connections are broken, and therefore a transient

nonreinforced condition exists. However, the symbols correctly indicate the status after the initial transient. In this connection it is again worth noting that while *single* occasional nonreinforced trials drive some delta cells to extinction, many have a sufficiently long gamma phase to be unaffected; therefore, the habit is given an increment of immunity to single nonreinforced trials. To avoid this obstruction to differentiation it is necessary to apply several successive nonreinforced trials (1, p. 170).

COMPLEX GENERALIZED NEURONS

Another paper (3) mentioned the need for more complex structures to cope with both laboratory and "natural" environments and gave a definition of a monopolar, generalized neuron suitable for use in elementary environments. Since the requirements for bipolar generalized neurons have not been studied thoroughly, no definition is offered.

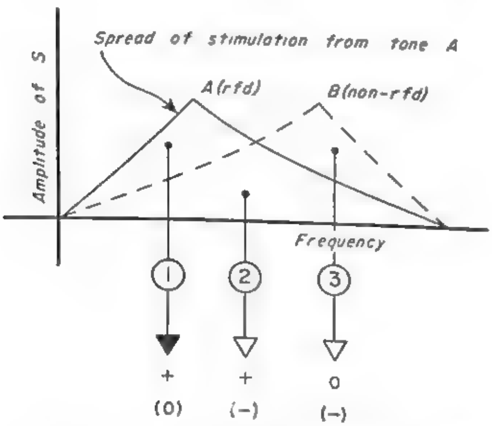


FIG. 4. When all the facts are not known, a number of possible explanations can exist for the same phenomenon. It is postulated in this figure that any auditory stimulus excites the entire auditory field. The variation in magnitude of stimulation serves to classify the neurons in a functional manner not contemplated in their structures. Neuron *2* is generalized and driven to extinction by the method of contrasts though it has precisely the same internal properties as neuron *1* which gains control of the response.

For the present paper it suffices to introduce the subject and indicate its significance.

Conditioning to tone *A*, say, shows generalization over most, if not all, of the audio spectrum (5, p. 113). If this result were attributed to direct stimulation of the entire basilar membrane, and if only monopolars are postulated, it is still possible to apply any tone without getting a response from all auditory neurons because of the gradient of stimulation and the distribution of thresholds as shown in Fig. 4. Then it is also possible to effect differentiation, because some neurons that are actuated on reinforced trials receive no stimulation on nonreinforced trials—these are stable and gain exclusive control of the motor cells in a laboratory environment. Galambos and Davis (4, p. 44) have shown that *receptors* are not excited by all tones, even though the receptors have low thresholds; but this does not rule out the possibility that "point" stimulation spreads to other neurons at subsequent stages in the nervous system and therefore acts like direct stimulation of the entire basilar membrane.

Neuron *2*, in Fig. 4, though structurally identical with the others, because of its location and threshold acts like a generalized neuron. In a natural environment it is rapidly driven to extinction—differentiation is effected—and *1*, having a somewhat higher threshold, survives longer but ultimately shares the same fate.

At least the gross aspects of generalization and differentiation are portrayed by several possible mechanisms and the author cannot as yet show an array of evidence that elects any one of the group. The subject needs much more study, especially so in view of the probability that more than one mechanism is in use. For practical reasons the fundamental properties originally postulated

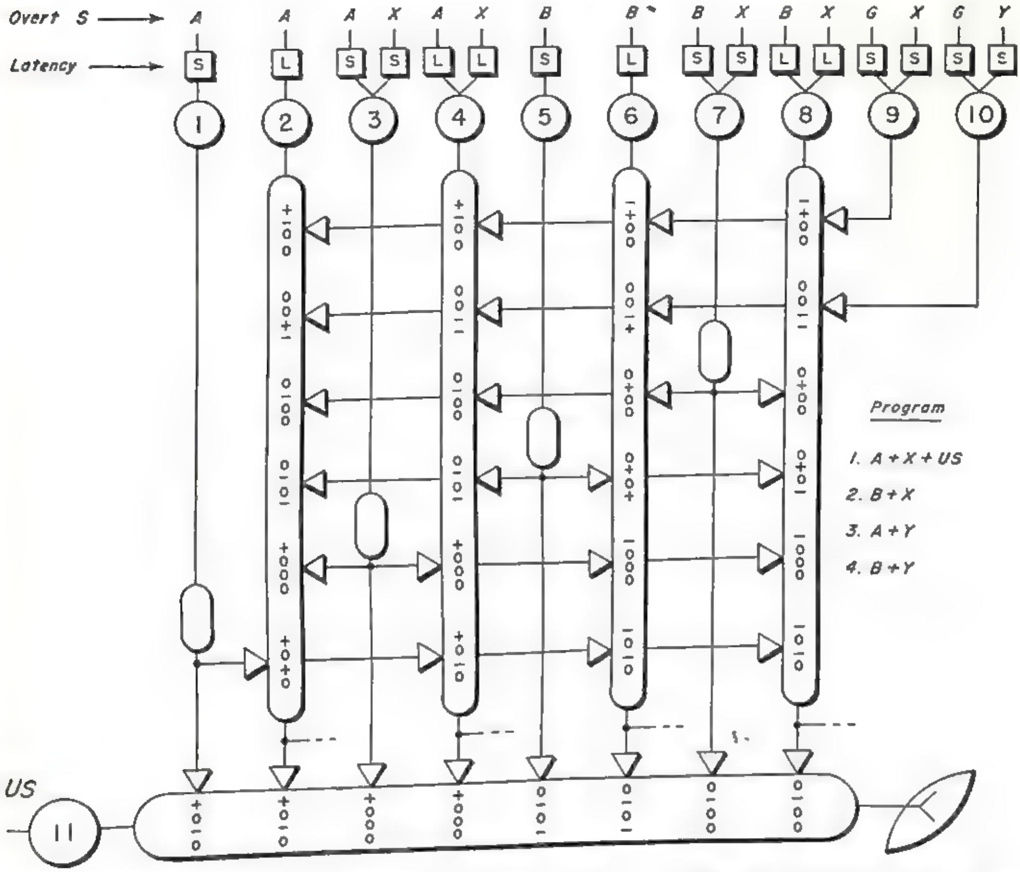


FIG. 5. The elements of the program are considered to be applied in random order. The first two represent the laboratory environment, the latter a "natural" environment. Neither the undifferentiated nor the differentiated laboratory-established CR is significantly affected by the natural environment provided the monopolar population is relatively small and the generalized neurons are functionally bipolar.

(1, p. 173) are retained and incorporated in a bipolar structure.

Figure 5 depicts the mechanism of generalization and differentiation anchored in bipolar neurons so that exposure to varied environments does not constitute cause for disagreement with fact. It will be observed that a generalized neuron now has its generalization property confined to one pole of a bipolar neuron so that its action is dependent on an additional factor in the environment. At the same time generalization, as a phenomenon, is not restricted because there are many generalized neurons with a broad range of properties to cover the action of all sensory fields. It is realized that many of these BA concepts are arbitrary, some possibly in the extreme, but they seem to be useful even now and it is hoped that by successive approximations they will in time bear a greater semblance to reality—whatever it may be.

CONCLUSION

Science displays numerous instances where mutually exclusive concepts are found to be limited in scope so that both points of view are finally merged in subsequent theories. Since the original concepts have some substantial basis in

fact, we do not expect subsequent theories to be fabricated from entirely foreign elements. We *do* expect them to unite in one system some of the elements which had previously been considered to be mutually incompatible. One such example is the AT hypothesis which demonstrates the probable coexistence of a neural mechanism which conditions only in the presence of conventional overt reinforcement and of another mechanism which denies the necessity for such reinforcement.

The AT's and the experimental data on which they are based again emphasize the concept that reinforcement is essentially no more than the coincidence of two neural events. It is believed that the tendency to consider various external agents as reinforcement is anthropomorphic in origin and is responsible for the fact that one external agent, food for example, is classified as reinforcement in various situations without sufficient concern for other factors (2, p. 458). The importance of considering the S situation *as viewed by the nervous system* cannot be overstressed.

The slight degree of genuine understanding that exists concerning the nature of intelligence is illustrated by the ease with which we can ask questions that embarrass a theory. For example: if sensory association exists, as postulated, why do we not see a flash of light after a particular S if the two had previously been associated? This question may have kept others from making a conjecture as to the nature of the mechanism. Possibly the author should follow suit, for it is only too apparent that either the hypothesis is inadequate or the point of view is incorrect.

These disquieting facts breed skepticism—and rightly so. However, the Brain Analogy should be viewed not as possessing any finality, but as an effective instrument for research, as a means of avoiding some of the inconsistencies of verbal arguments, and as a foundation for the establishment of the interchangeable constants that characterize the more mature sciences. If the present paper does no more than suggest significant lines for experimental attack, it is worth the effort.

Summary

The phenomenon of transfer of stimulus differentiation is explained in terms of the Brain Analogy theory. It is shown to be a property of the same structure, the association tracts, which accounts for the phenomena of secondary conditioning, conditioned inhibition, and sensory preconditioning.

REFERENCES

1. Coburn, H. E. The brain analogy. *Psychol. Rev.*, 1951, 58, 155–178.
2. Coburn, H. E. The brain analogy: a discussion. *Psychol. Rev.*, 1952, 59, 453–460.
3. Coburn, H. E. The brain analogy: association tracts. *Psychol. Rev.*, 1953, 60, 197–206.
4. Galambos, R., & Davis, H. The response of single auditory-nerve fibers to acoustic stimulation. *J. Neurophysiol.*, 1943, 6, 39–57.
5. Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep) London: Oxford Univer. Press, 1927.

[MS. received January 15, 1953]

Psychological Review
Vol. 60, No. 6, 1953

A COMMENT ON BURKE'S ADDITIVE SCALES AND STATISTICS

VIRGINIA L. SENDERS

Antioch College

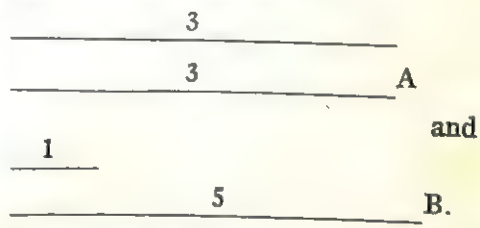
Burke, in a recent article (1), states that "statistical technique begins and ends with numbers and with statements about them." Therefore, he concludes, "the properties of a set of numbers *as a measurement scale* should have no effect upon the choice of statistical techniques for representing and interpreting the numbers." Again, referring to statistical interpretation, Burke states that "the use of the sample mean and standard deviation does no violence upon the data, *whatever the properties of the measurement scale.* Thus, the use of the usual statistical tests is limited only by the well-known statistical restrictions" (italics mine).

Some rather unfortunate implications follow from Burke's position. When numbers have been assigned to objects according to some stated rules these numbers can indeed be manipulated in any way we desire. But once the manipulations have been completed and the tests made, a dilemma arises. If we have performed operations on the numbers which we could not perform on the objects, we must choose between two interpretive procedures. We must either assume that (*a*) what is true of the numbers is also true of the objects, or (*b*) what is true of the numbers is not necessarily true of the objects.

The first assumption leads to all sorts of difficulties, which have been adequately described by Campbell (2), Stevens (4), Reese (3), and others. These difficulties may be illustrated, in oversimplified form, by two examples, one numerical and one geometric.

Suppose we have a measurement scale which has ordinal but not interval or additive properties. Such a yardstick is illustrated below:

An object is measured by laying it against this yardstick in the usual way. We are given two pairs of objects, A and B, whose lengths are:



The numbers assigned to the objects total 6 in both cases, but inspection will reveal that the additions of the objects will give a longer line in case A than in case B.

Thus the statement that *there is no difference between the sums of the numbers used to represent length in the two cases* is correct, but strongly suggests the erroneous conclusion that if the two summed lines were laid side by side, no difference could be discerned between them.

The absurdity of possible conclusions may also be revealed by a numerical example. Suppose families are grouped according to income on a scale where the number 5 means "very rich" and number 1, "very poor." The actual income intervals, however, are unequal, as follows:

Number	*Income limits*	*Midpoint*
5	$5,000–$1,000,000	$502,500
4	3,000– 5,000	4,000
3	2,000– 3,000	2,500
2	1,000– 2,000	1,500
1	0– 1,000	500

In town A all the families have incomes which fall in class 3, while in town B half the families have incomes in class 5, and half, incomes in class 1. Both towns will have a *mean* of 3 if we consider only the numbers assigned to the categories, but it is evident that town B, with a mean income of $251,000 is richer than town A with a mean income of $2,500.

If we accept conclusion *b*, on the other hand, we are in an even more ridiculous position. Though our statistical procedures may have been perfectly justified and our interpretations correct when considered strictly in relation to the *numbers*, we can make no interpretation about the properties of the objects to which the numbers have been assigned. As psychologists, we can draw no conclusions about responses, organisms, or behavior, but only about numbers.

Since psychologists are presumably more interested in the behavior they describe with numbers than in the numbers themselves, they will learn more if their statistical techniques correspond with the properties of the set of numbers as a measurement scale than if these properties "have no effect upon the choice of statistical techniques for representing and interpreting the numbers."

REFERENCES

1. BURKE, C. J. Additive scales and statistics. *Psychol. Rev.*, 1953, **60**, 73–75.
2. CAMPBELL, N. R., *et al.* Final report. *Advanc. Sci.*, 1940, No. 2, 331–349.
3. REESE, T. W. The application of the theory of physical measurement to the measurement of psychological magnitudes, with three experimental examples. *Psychol. Monogr.*, 1943, **55**, No. 3 (Whole No. 251).
4. STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1–49.

[MS. received June 8, 1953]